An In-Silico Based Multi Epitope Vaccine Construction of Flagellin Protein Opposing Helicobacter Pylori Bacteria (*H. Pylori*)

By

Tahamina Akhter
19346005

A thesis submitted to the School of Pharmacy in partial fulfillment of the requirements for the degree of Bachelor of Pharmacy (Hons.)

School of Pharmacy
Brac University
February 2024

Declaration

It is hereby declared that

1. The thesis submitted is my own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I have acknowledged all main sources of help.

Student's Full Name & Signature:

Tahamina Akhter
ID: 19346005

Approval

The thesis titled "An In-Silico Based Multi Epitope Vaccine Construction of Flagellin Protein Opposing Helicobacter Pylori Bacteria (*H. Pylori*)" submitted by Tahamina Akhter (19346005) of Summer, 2019 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy.

Supervisor:

_____

Mohammad Kawsar Sharif Siam
Senior Lecturer
School of Pharmacy
BRAC University

Program Coordinator:

_____

Dr. Hasina Yasmin
Professor, Assistant Dean and Program Director
School of Pharmacy
Brac University

Dean:

_____

Eva Rahman Kabir, PhD
Professor and Dean,
School of Pharmacy
Brac University

Ethics Statement

Thesis was done within maintaining all ethical standards. Any unethical activities were avoided. No human or animal trials were operated during research.

Abstract

Helicobacter pylori (*H. pylori*) bacteria is a gram-negative bacteria that has capability to cause serious disease in human stomach by leading into peptic ulcer, mucosal lymph tissue, stomach cancer. According to WHO, it became type I carcinogen bacteria which colonizes and affects gastric mucosa layer of the stomach. Available antibiotics drugs were effective but now-days growing resistance to this particular bacterium. Hence, for such emerging situation a vaccine will be much more effective against bacterium; more convenient and budget friendly option to sufferer. In this study, protein sequence is used with different epitopes of Helper T cells, cytotoxic T cells, B cells to predict multi epitope-based vaccine for protection by computational (immunoinformatics) method.

Keywords: FlaA protein; *H. pylori*; multi-epitope vaccine; in-silico; CTL; HTL

Acknowledgement

This work was assisted and supervised by my supervisor, Mohammad Kawsar Sharif Siam, Senior lecturer, School of Pharmacy, Brac University, to whom I would like to express my gratitude for supporting, guiding and inspiring me.

Tahamina Akhter

ID: 19346005

School of Pharmacy

Brac University

Table of Contents

List of Figures

List of Tables

List of Acronyms

CTL	Cytotoxic T Lymphocyte
HTL	Helper T Lymphocyte
BabA	Blood group antigen binding adhesion
SabA	Sialic acid binding adhesion
VacA	Vacoulating cytotoxin
RMSD	Root mean square deviation

Chapter 1

Introduction

Helicobacter pylori (*H. pylori*) is one of the leading bacteria that mainly associated with stomach infections; several diseases peptic ulcer, stomach cancer, mucosal lymph tissue (Chehelgerdi et al., 2023). In the worldwide, near 44.5% people are affected through *H. pylori* (Sabbagh et al., 2019). WHO (World health organization) in 1994 declared *H. pylori* as a group-I-carcinogen. Infectious peoples with *H. pylori* are in risk of building up peptic ulcer, gastric cancer; peptic ulcer can see in 10%-50% infected patients and among them 1% to 3% lead to gastric cancer. *H. pylori* increases 2 to 7 times risk of gastric cancer compare to unaffected one (Y. C. Wang, 2014). They also related with other diseases e.g. iron deficiency, vitamin B12 deficiency, thrombocytopenic purpura etc (Ding, 2020). Prevalence rate is more than 80% in the developing countries and 40% in the developed countries. Most of time, it develops in the childhood and can see throughout in the rest of life unless any treatment is given. *H. pylori* transmitted from environment; water, contaminated vegetables. Also, transmission can be occurred from person to person through fecal-oral, oral-oral, gastro-oral routes (Sabbagh et al., 2019).

Less number of antibacterial agents are effective against *H. pylori* infections. As few numbers of antibiotics can be used, it rapidly showing primary antibiotic resistance. First-line (a proton pump inhibitor with two antibiotics) and rescue therapies in past decades, showing decline in around 10%-30% patients worldwide. WHO catalogued *H. pylori* as most serious among 20 pathogens since 2017; due to their drug resistance they causing emerging threat to human health (Tshibangu-Kabamba & Yamaoka, 2021).

Vaccine will be a great option to deal with antibiotic resistance situation. Traditional vaccine develop method requires more time and expenses rather than subunit vaccine (Ma et al., 2021).

Immunoinformatics method uses computational and immnunological tools to develop vaccine that are more time convenience and less expensive (Kar et al., 2022). Immunoformatics tools detect T-cells epitope that will enhance immune activity which will be using in vaccine design. Along with, vaccine efficacy, safety, antigenicity etc. can be checked out through tools (De Groot et al., 2020).

However, *H. pylori* flagella consist of virulence factors, so it can be used as a vaccine target (Kao et al., 2016).

1.1: *H. pylori* structure, genome and functional aspects

H. pylori is a spiral shaped, gram-negative, flagellated bacterium that be included in class Epsilonproteobacteria, phylum Proteobacteria, family Helicobacteraceae, order Campylobacterales, genus Helicobacter (Kira & Isobe, 2019). Moreover, three fully sequenced genome strains have been found yet and they are 26695, J99, HPAG1 strains. Their base pair numbers are 1643831 for J99 strains, 1667867 for 26695 strains and 1596366 for HPAG1 (Alm et al., 1999; *HELICOBACTER PYLORI - Biological Agents - NCBI Bookshelf*, n.d.; Oh et al., 2006). The strains have protein coding genes respectively 1495, 1552, 1536 of J99, 26695, HPAG1 (*HELICOBACTER PYLORI - Biological Agents - NCBI Bookshelf*, n.d.). 26695-strain have IS605 sequence, 5SRNA in one end and 521 base pair in another end of two regions (region 1,3) among five regions of G+C compositions. For DNA processing, these two regions provide genes. This strain has 36 species of tRNA: two 23S-5S, two 16S, one 5S of rRNA and two insertion sequence of DNA IS605 with 5 full lengths of 13 copies, IS606 with 2 full lengths of 4 copies. Bacterial chaperones genes GroES, DnaK, CbpA, GrpE, DnaJ, GroEL, and HtpG have role in transcription and creates difference from E. coli mechanism. Although, similarities can be seen in cell division, cell replication and in secretion (Tomb et al., 1997).

Furthermore, in-vitro experiment recognized almost 43 outer membrane proteins, 48 enzymes protein, 11 flagella components, 9 binding and transport proteins, 8 cytotoxic associated genes pathogenicity island (Zanotti & Cendron, 2014).

The bacterium has 0.5 to 1.0 μm width, 2.4 to 4.0 μm length and some unipolar flagella with 2-6 of characteristics (Goodwin et al., 1989; *HELICOBACTER PYLORI - Biological Agents - NCBI Bookshelf*, n.d.). The flagella that help in motility, have 30 nm thickness with 2.5 μm length (*HELICOBACTER PYLORI - Biological Agents - NCBI Bookshelf*, n.d.). More than 50 proteins were identified into flagellum that maintains structure or regulation. FlaA, FlaB are two subunits of filament of flagella that help in virulence and colonization (Van Amsterdam et al., 2006). FlaA is a surface structure and dominant protein of flagella. In J99 strains, among 510 amino acids of FlaA protein 214-353 part length was recognized as antigenic element that appropriate for vaccine designing (Zarei et al., 2017).

1.2 Pathogenesis of *H. pylori*

For pathogenesis and colonization four steps are essential for *H. pylori*; (i) Surviving: remain alive under stomach acidic environment; (ii) Motility: Motility and chemotaxis through flagella to epithelium cells; (iii) Ahesion: Adhere to host receptors; (iv) Toxin release: Toxin release that causes tissue damage.

H. pylori maintain urease activity that adjust periplasmic pH by which they can survive on acidic conditions and adjust phagosome pH, megasome formation by which macrophages can be avoided. Urel (Proton gated urea channel), modulates urease activity that deal with entry of urea to obstruct lethal alkalization. When gates are open at pH 5.0 huge amount of urea appear in the bacterium that produce ammonium. Ammonium may can allow protein rapid neutralization. Ammonium hydroxide produces by break down of urea to ammonia and carbon-

dioxide by extracellular urease that neutralize acid condition and make suitable for bacteria (Kao et al., 2016).

Flagella of *H. pylori* assist to reach mucus layer of gastric mucosa. LuxS enzyme that aid in motility, catalyzes producing of autoinducer 2 (AI- 2) that needed in Fur (ferric uptake regulator) which is also essential for motility (Camilo et al., 2017). Basal body of flagella provide motility energy. Higher motility is linked with end results of critical pathology as they increase density of bacteria, high response of inflammation in the upper stomach (Kao et al., 2016). Through four chemoreceptors (Tlp A, B, C, D) they sense chemical environment and make role in chemotaxis (Camilo et al., 2017; Cid et al., 2013). In the infected and inflamed antrum, Tlp D assist in surviving and growing (Cid et al., 2013). A study showed that Tlp A and Tlp D inactivation in mice lead to gastric colonization reduction (Camilo et al., 2017).

BabA, SabA, OipA, HopQ, AlpA, AlpB are some of the outer-membrane proteins (OMP). BabA, HopQ proteins increase translocation of CagA gene to adhere to host cells through T4SS; SabA protein also help in adhesion to host cells; OipA protein give a rise of production of inflammatory cytokines; and AlpA, AlpB proteins help in adhesion to extracellular matrix (Matsuo et al., 2017). BabA that has similar structure of O blood type antigen attach to fucosylated Lewis B blood group antigen (Le^b) that expresses on the gastric epithelial cells (Camilo et al., 2017; Kao et al., 2016).

CagA has virulence factor that administer into host cell through cagPAI which encrypt T4SS (Type-4 secretion system). Probably CagA is larger than the gates of T4SS. So, they use beta integrin receptor to deliver CagA into cells. After administration, CagA associates with phosphatidylserine and connect with cell membrane's inner leaflet. When it administered into cytoplasm, CagA itself being phosphorylated and make changes in host cell signal of phosphorylation independent, phosphorylation dependent manner. To the phosphatase SHP-2,

phosphorylated CagA binds with and throw effects on cell's adhesion, migration, spreading (Cid et al., 2013; Kao et al., 2016). Moreover, VacA are relates with increased mitogen activated with protein kinase, intrinsic apoptosis, autophagy, changes in immune response, cell death. P^{33}, P^{55} are domains of secreted VacA that have oligomeric structure and anion selection characteristics channel via they secrete anion, bio-carbonate into the host of cytoplasm. Probably, this channel also provides metabolic components for bacterial growth (Cid et al., 2013).

Chapter 2

Materials and Method

A flowchart showing methods of vaccine prediction and vaccine validation in figure 1.

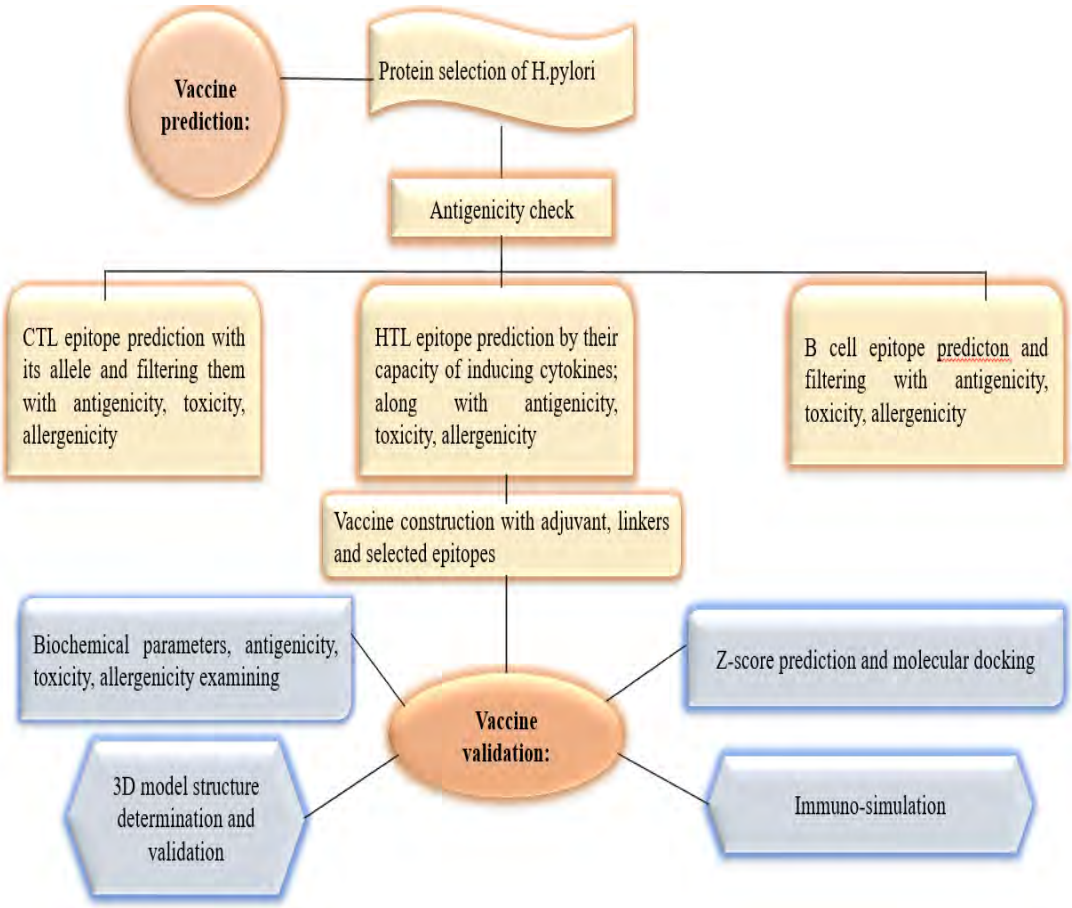


Figure 1: Procedure of building multi epitope vaccine for H. pylori

2.1 Retrieval of Protein Sequence

Trough Uniport database, appropriate protein was selected that showed antigenicity. The selected target protein was FlaA (Flagellin A) that retrieved with FASTA format from uniport database (https://www.uniprot.org/).

Antigenicity was determined by using Vaxijen v2.0 server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) and target organism was chosen as bacteria; threshold was set at 0.5 (Doytchinova & Flower, 2007). As the protein showed good amino acid length and quality, it was selected for vaccine development.

2.2 Cytotoxic T cell Lymphocyte (CTL) epitope and Allele selection

CTL find differences of healthy cells and infected cells (M. V. Larsen et al., 2005). CTL control diseases by excluding pathogens or cancer cells through secreting various cytokines (Ito & Seishima, 2010).When CTL epitopes bind to MHC class-I, CTL able to recognize them and trigger immune response (M. V. Larsen et al., 2005). For CTL epitope prediction, Net CTL 1.2 server (https://services.healthtech.dtu.dk/services/NetCTL-1.2/) was used as it has good prediction abilities. They identified CTL epitope that has Tap transport abilities for peptide transportation into endoplasmic reticulum (M. V. Larsen et al., 2007). To use server for prediction, all the parameters were kept as default but sorting score were set into combined score. Protein sequence were inputted as FASTA format in the blank box.

To evaluated CTL epitope, strongly binding alleles to MHC class-I were selected through (https://services.healthtech.dtu.dk/services/NetMHCIIpan-4.1/) (Reynisson et al., 2020). All the CTL epitopes with FASTA format inputted into blank box and length were selected as 9mer peptide, all the alleles were selected with BA prediction.

Furthermore, antigenicity, allergenicity, toxicity were determined for alleles. For antigenicity (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) server were used with 0.5 threshold and bacteria target organism (Doytchinova & Flower, 2007). For allergenicity (https://www.ddg-pharmfac.net/AllerTOP/) were used with one by one CTL epitopes (Dimitrov et al., 2013). For toxicity (https://webs.iiitd.edu.in/raghava/toxinpred/) server were used and batch submission were selected (Gupta et al., 2013).

2.3 HTL (Helper T-cell Lymphocytes) epitope selection

HTL induces response of cellular and humoral immune systems by induction of CTL and antibody responses. Along with, they secrete different lymphokines against bacteria, parasite, virus etc. When antigen bind with MHC class-II, they able to recognize them (Alexander et al., 1998). For HTL epitope determination, (https://services.healthtech.dtu.dk/service.php?NetMHCIIpan-4.0) server were used (Montes-Grajales & Olivero-Verbe, 2021). To use the server peptide length were set as 15 mer, protein sequence put on given box, alleles were selected with BA prediction.

For HTL epitopes, antigenicity, allergenicity, toxicity were determined with the same way that was used in CTL selection.

2.4 HTL capability to induce cytokines

IFN-γ produced by HTL cells that can control bacterial infection (Bao et al., 2014). Also, they regulate response of cell mediate immune system, activation of macrophage and work with APC (Antigen Presenting Cell) (Walker et al., 2021). For checking IFN-γ inducing capabilities (https://webs.iiitd.edu.in/raghava/ifnepitope/predict.php) server were operated (Kalkal et al., 2022).

IL-4 are important for B cells surviving and growing. Along with, they switch immunoglobulin to IgG1 and IgE. For checking IL-4 epitope inducing capabilities (https://webs.iiitd.edu.in/raghava/il4pred/design.php) server were operated (None et al., 2021) and all the modes were kept in default system.

IL-10 derived from HTL and have a role in preventing autoimmune and inflammatory diseases, maintain homeostasis during tissue injury and acute type infections (Iyer & Cheng, 2012).

(https://webs.iiitd.edu.in/raghava/il10pred/predict3.php) server were operated during capability checking of HTL (Kalkal et al., 2022) and all the modes were kept in default system.

2.5 B-cell Epitope Selection

B cells supply prolonged immune protection by producing antibody. Linear epitopes of B cells are linear expanses of antigen protein residues. For selection of linear epitopes of B cells (http://tools.iedb.org/bcell/) server were used. Protein sequence were submitted in plain format with bepipred linear epitope prediction selection (Jespersen et al., 2017).

Antigenicity, allergenicity, toxicity determined for B-cells epitopes with the same process that used for CTL, HTL epitopes.

2.6 Vaccine Construction with Linkers

Linkers are short form of amino acids. Without using linkers, the vaccine result will be undesirable, with low yields results and misfolding can occur (Gong et al., 2022). For constructing multi-epitope vaccine, different linkers were used to connect different epitopes and adjuvant together. EAAAK linker joined adjuvant to the epitopes, AAY linker joined CTL, GPGPG linker joined HTL, KK linker joined B-cells together (Ayyagari et al., 2022). These are rigid linkers that give proper distance to proteins and different epitopes so that interaction can be reduce and their biological activity will be in protection (Gong et al., 2022).

2.7 Vaccine Antigenicity, Toxicity, Allergenicity Evaluation

Antigenicity again checked for final constructed vaccine with following sever (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) and parameters were set as bacteria organism, 0.5 threshold (Doytchinova & Flower, 2007).

To checking toxicity for vaccine (http://www.t3db.ca/) server were used with sequence search option (Gong et al., 2022). Vaccine sequence inputted in FASTA format and all the parameters were kept in default mode.

To checking allergenicity for vaccine sequence (http://www.allergenonline.org/) server were used with sequence search and in FASTA format (Sircar et al., 2014).

2.8 Vaccine's Biochemical Analysis

Physiochemical properties such as GRAVY, molecular weight, therapeutic pI, aliphatic index, instability index, estimated half-life etc. prediction was checked through (https://web.expasy.org/protparam/) server (Ayyagari et al., 2022). Half life estimation result is provided for human, E. coli, yeast organism by observing protein sequence's N-terminal amino acids that indicates time for fading half amount of protein in the cell during synthesis. Instability index indicates stability in the test tube of vaccine or protein and rate should be in below 40. GRAVY (Grand Average of Hydropathy) indicates hydropathy value sum of all amino acid that divided with residue number (Gasteiger et al., n.d.). Negative value of GRAVY imply hydrophilicity and positive value indicates hydrophobicity. Aliphatic index regards to aliphatic side chain volume (H. Wang et al., 2021). Extinction co-efficient is about light absorbing of protein in a particular wavelength (Gasteiger et al., n.d.).

2.9 3D model generation of constructed vaccine

For 3-dimensional model generation Phyr^2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) server were being used in intensive mode with single sequence and they provide model by folding simulation of simplified ab initio and multiple template modelling combination (Kelley et al., 2015).

2.10 Validation of 3D model

For 3-diemnsional model validation (https://swissmodel.expasy.org/) sever were being operated that provide details of homology models such as global and local quality, ligands, Ramachandran plot etc. Mainly Ramachandran plot envisions about amino acid residue of protein or vaccine structure (Waterhouse et al., 2018).

2.11 Quality Justification of 3D model

To justify 3D model quality, z-score and error checking, ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php) server is being used. Server measure score of overall quality and show in plot that indicates score of experimented available protein chain in Protein Data Bank (Wiederstein & Sippl, 2007).

In addition, z-score is derived from server results show overall model quality and give measurement how far the total energy of structure deviates from actual result from a random confirmation distribution (Wiederstein & Sippl, 2007).

2.12 Molecular Docking

Molecular docking is the process of visualize binding affinity and interaction among constructed vaccine and human toll like receptor. Human toll like receptor 5 (TLR5) was retrieved from RCSB protein data bank in PDB format (PDB ID: 3J0A). In accordance with Israel T. Desta, the molecular docking was carried out with ClusPro server (https://cluspro.org/home.php). The server is based upon fast fourier transform (FFT) which is known as PIPER that deposit one protein on fixed-grid and another protein on movable-grid. The server let cover gap between receptor and ligand with adjustment to simulate binding process (Desta et al., 2020).

2.13 Immune Simulation of Vaccine

In the last stage, immune response prediction was done with C-immsim server (https://kraken.iac.rm.cnr.it/C-IMMSIM/) to evaluate vaccine whether it can induce immunogenicity in the body or not (Rapin et al., 2010). The server represents different immune responses of humoral and cellular immune systems such as B lymphocytes, T lymphocytes, innate immune cells etc. in graphical forms (Rueckert & Guzmán, 2012). In the server, three injection doses were added with vaccine sequence by 1, 84, 168time steps. One time step is equal to 8hours time and interval between two doses are 4weeks (Nain et al., 2020).

2.14 Assertion of materials and methods

Helicobacter pylori vaccine design, prediction and validation were performed through in-silico method. Different tools of online servers were essential keys for developing vaccine that thoroughly using in different literatures. Moreover, the aim was to predict a useful therapeutic vaccine for human kind but uncertain issue can come over; so further inspection is needed.

Chapter 3

Results

3.1 Protein collection and antigenicity

Structural proteins were screened by their amino acid number and antigenicity. After that Flagellin A protein (Uniport ID: P0A0S1) was downloaded with FASTA format from Uniport protein data base. The full sequence of protein and feature (Table 1) is given lower:

Table 1: Feature of Protein (Bateman et al., 2023)

Protein name	**Flagellin A**
Gene	flaA
Amino acids	510
Organism	Helicobacter pylori (strain ATCC 700392/26695) (Campylobacter pylori)
Status	UniportKB reviewed (Swiss-Prot)
Last updated	2007-01-23 v2
Mass	53,284

Full sequence:

```
MAFQVNTNINAMNAHVQSALTQNALKTSLER
LSSGLRINKAADDASGMTVADSLRSQASSLG
QAIANTNDGMGIIQVADKAMDEQLKILDTVK
VKATQAAQDGQTTESRKAIQSDIVRLIQGLD
NIGNTTTYNGQALLSGQFTNKEFQVGAYSNQ
SIKASIGSTTSDKIGQVRIATGALITASGDI
```

SLTFKQVDGVNDVTLESVKVSSSAGTGIGVL
AEVINKNSNRTGVKAYASVITTSDVAVQSGS
LSNLTLNGIHLGNIADIKKNDSDGRLVAAIN
AVTSETGVEAYTDQKGRLNLRSIDGRGIEIK
TDSVSNGPSALTMVNGGQDLTKGSTNYGRLS
LTRLDAKSINVVSASDSQHLGFTAIGFGESQ
VAETTVNLRDVTGNFNANVKSASGANYNAVI
ASGNQSLGSGVTTLRGAMVVIDIAESAMKML
DKVRSDLGSVQNQMISTVNNISITQVNVKAA
ESQIRDVDFAEESANFNKNNILAQSGSYAMS
QANTVQQNILRLLT

To add, antigenicity was determined through Vaxijen v2.0 software and gave result of 0.7748 (Figure 2).

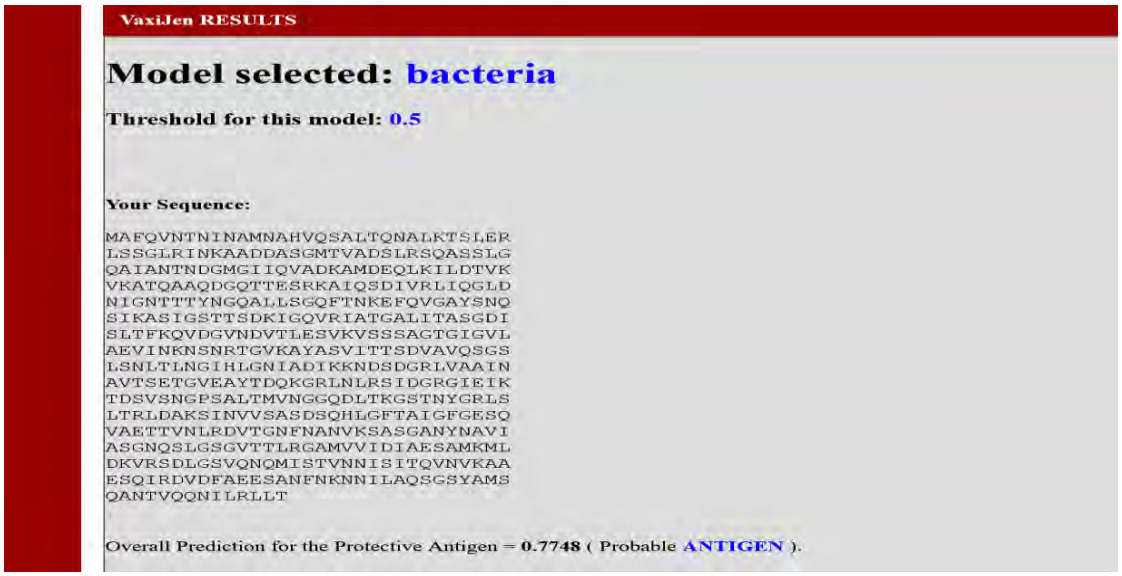


Figure 2: Antigenicity score of protein in Vaxijen v2.0 server (Doytchinova & Flower, 2007)

3.2 CTL and Allele Screening:

NetCTL1.2 software helped to find out CTL epitopes with 0.75 threshold, A1 supertype, 0.05 Tap transport efficiency and combined score (M. V. Larsen et al., 2007). 9 predicted CTL epitopes were being showed by software (Figure 3). Epitopes that have <-E were being selected for first stage.

```
NetCTL-1.2 predictions using MHC supertype A1. Threshold 0.750000

 282 ID FASTA pep TSETGVEAY aff   0.7313 aff_rescale   3.1050 cle 0.7686 tap   2.8470 COMB   3.3626 <-E
 355 ID FASTA pep ASDSQHLGF aff   0.5127 aff_rescale   2.1770 cle 0.6611 tap   2.5970 COMB   2.4060 <-E
 113 ID FASTA pep QSDIVRLIQ aff   0.2457 aff_rescale   1.0433 cle 0.0576 tap  -0.2190 COMB   1.0410 <-E
 182 ID FASTA pep ASGDISLTF aff   0.1654 aff_rescale   0.7023 cle 0.9572 tap   2.5660 COMB   0.9742 <-E
 290 ID FASTA pep YTDQKGRLN aff   0.2219 aff_rescale   0.9422 cle 0.0288 tap  -1.3790 COMB   0.8776 <-E
 485 ID FASTA pep NILAQSGSY aff   0.1317 aff_rescale   0.5592 cle 0.9636 tap   3.2490 COMB   0.8662 <-E
 329 ID FASTA pep DLTKGSTNY aff   0.1378 aff_rescale   0.5851 cle 0.9729 tap   2.5430 COMB   0.8582 <-E
 164 ID FASTA pep TTSDKIGQV aff   0.1588 aff_rescale   0.6743 cle 0.9427 tap   0.3220 COMB   0.8318 <-E
 144 ID FASTA pep NKEFQVGAY aff   0.1316 aff_rescale   0.5586 cle 0.8410 tap   2.8770 COMB   0.8286 <-E
 433 ID FASTA pep MLDKVRSDL aff   0.1159 aff_rescale   0.4923 cle 0.9621 tap   0.9460 COMB   0.6839
 344 ID FASTA pep RLDAKSINV aff   0.1211 aff_rescale   0.5141 cle 0.9612 tap   0.4120 COMB   0.6789
  17 ID FASTA pep QSALTQNAL aff   0.1134 aff_rescale   0.4813 cle 0.9504 tap   1.0720 COMB   0.6774
 391 ID FASTA pep VKSASGANY aff   0.0952 aff_rescale   0.4042 cle 0.6823 tap   3.1730 COMB   0.6651
 301 ID FASTA pep SIDGRGIEI aff   0.1143 aff_rescale   0.4854 cle 0.9607 tap   0.6720 COMB   0.6631
 383 ID FASTA pep VTGNFNANV aff   0.1219 aff_rescale   0.5175 cle 0.9180 tap   0.0530 COMB   0.6579
  67 ID FASTA pep NTNDGMGII aff   0.1269 aff_rescale   0.5386 cle 0.5362 tap   0.6100 COMB   0.6496
 128 ID FASTA pep NTTTYNGQA aff   0.1337 aff_rescale   0.5676 cle 0.6593 tap  -0.5860 COMB   0.6372
 334 ID FASTA pep STNYGRLSL aff   0.1003 aff_rescale   0.4257 cle 0.8421 tap   0.8620 COMB   0.5951
 180 ID FASTA pep ITASGDISL aff   0.0925 aff_rescale   0.3928 cle 0.9647 tap   1.0210 COMB   0.5886
 353 ID FASTA pep VSASDSQHL aff   0.0925 aff_rescale   0.3925 cle 0.8997 tap   1.1210 COMB   0.5835
 130 ID FASTA pep TTYNGQALL aff   0.0895 aff_rescale   0.3802 cle 0.9590 tap   1.1090 COMB   0.5795
 493 ID FASTA pep YAMSQANTV aff   0.1109 aff_rescale   0.4710 cle 0.5653 tap   0.4660 COMB   0.5791
 165 ID FASTA pep TSDKIGQVR aff   0.1130 aff_rescale   0.4796 cle 0.1804 tap   1.3380 COMB   0.5736
 500 ID FASTA pep TVQQNILRL aff   0.0834 aff_rescale   0.3543 cle 0.9675 tap   1.1810 COMB   0.5585
 379 ID FASTA pep NLRDVTGNF aff   0.0662 aff_rescale   0.2813 cle 0.9520 tap   2.6870 COMB   0.5584
 360 ID FASTA pep HLGFTAIGF aff   0.0721 aff_rescale   0.3060 cle 0.9208 tap   2.2510 COMB   0.5567
 428 ID FASTA pep ESAMKMLDK aff   0.1149 aff_rescale   0.4879 cle 0.2019 tap   0.4720 COMB   0.5418
 129 ID FASTA pep TTTYNGQAL aff   0.0919 aff_rescale   0.3900 cle 0.6255 tap   1.0170 COMB   0.5347
 124 ID FASTA pep DNIGNTTTY aff   0.0599 aff_rescale   0.2542 cle 0.9696 tap   2.5730 COMB   0.5283
 310 ID FASTA pep KTDSVSNGP aff   0.1206 aff_rescale   0.5120 cle 0.0323 tap   0.0230 COMB   0.5180
  28 ID FASTA pep SLERLSSGL aff   0.0774 aff_rescale   0.3286 cle 0.9481 tap   0.8930 COMB   0.5154
```

Figure 3: CTL prediction score with NetCTL1.2 software (M. V. Larsen et al., 2007)

Epitopes that got from figure 3 screened further through their combined score with > 0.7 and selected into second stage. The screened CTL with their combined score showed in Table 2 below:

Table 2:CTL epitopes with their combined score

CTL epitopes	Combined Score
TSETGVEAY	2.8470
ASDSQHLGF	2.5970
ASGDISLTF	2.5660
NILAQSGSY	3.2490
DLTKGSTNY	2.5430
NKEFQVGAY	2.8770

3.3 Allele screening for CTL epitopes

Screened CTL epitopes that were selected for 2nd stage, run into NetMHCPan4.1 server to find out their suitable binding alleles and result are shown in figure 4 (Reynisson et al., 2020). Among 6 epitopes 5 epitopes with binding allele were found that shown in table 3.

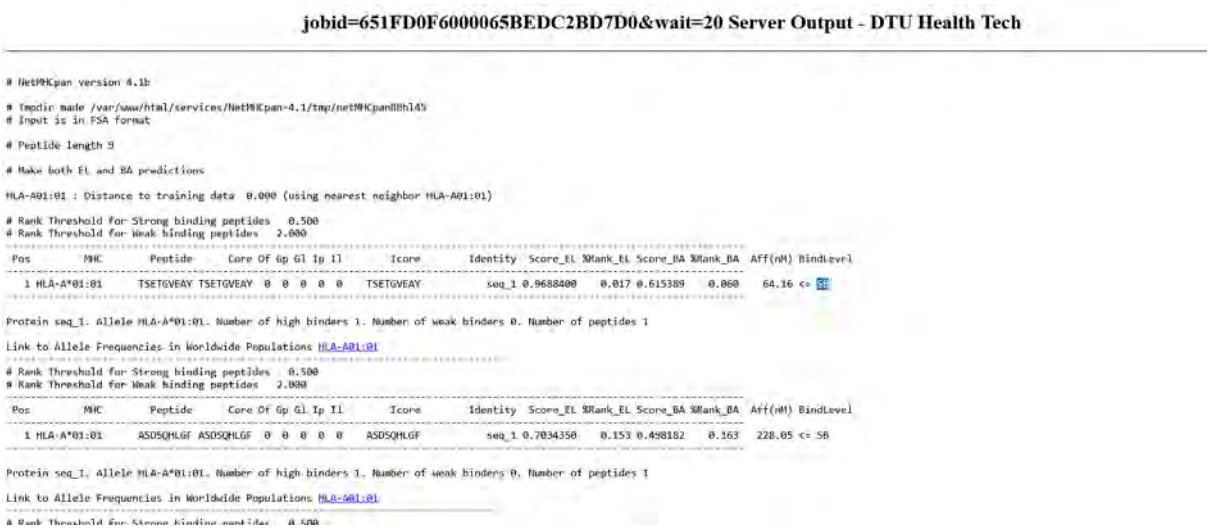


Figure 4:Strong Binding alleles prediction for MHC class-I on NETMHCpan 4.1 server (Reynisson et al., 2020)

Table 3:CTL epitopes with their MHC class-I alleles and ranks

Allele	Epitope	% Rank EL	% Rank BL
HLA-A*01:01	TSETGVEAY	0.017	0.060
HLA-A*01:01	ASDSQHLGF	0.153	0.163
HLA-A*26:01	NILAQSGSY	0.347	0.328
HLA-A*26:01	DLTKGSTNY	0.191	0.524
HLA-B*58:01	ASGDISLTF	0.120	0.400

3.4 Final CTL epitopes selection by antigenicity, allergenicity, toxicity

Epitopes with binding alleles goes into further screening process according to their antigenicity, allergenicity, toxicity. Vaxijen v2.0 for antigenicity (Doytchinova & Flower, 2007), AllerTOP v2.0 (Dimitrov et al., 2013) server for allergenicty, Toxinpred (Gupta et al., 2013) server for toxicity determination were being used. Toxinpred determined toxicity by their physiochemical properties such as molecular weight, hydrophobicity, hydrophilicity, CTL charge, hydropathicity. The results are showed in figure 5. Final CTL epitope selection were based onto them and 1 suitable CTL epitope was found (table 4).

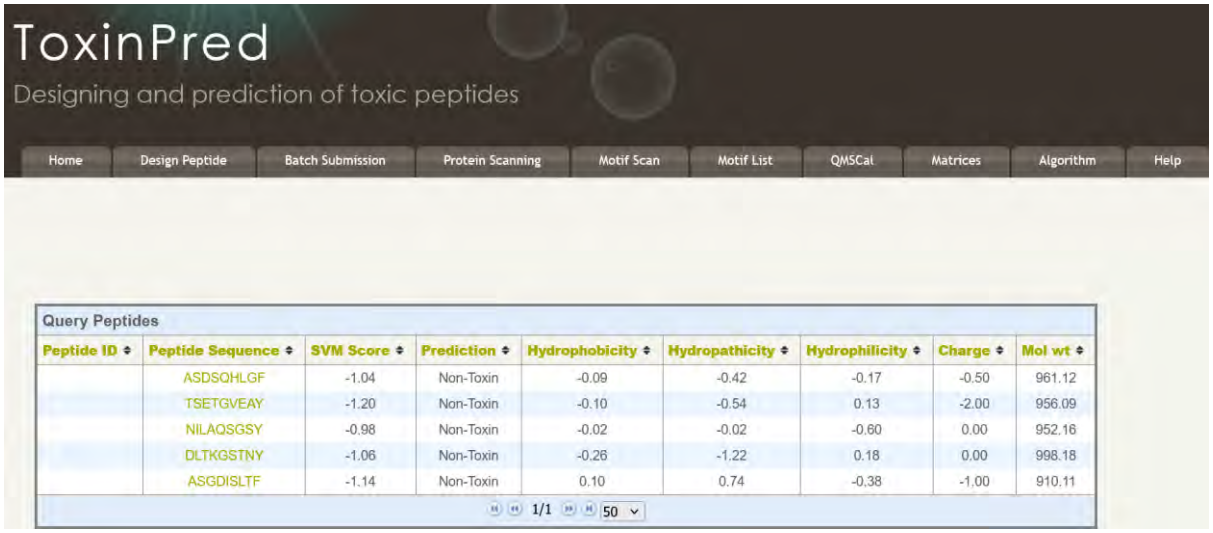


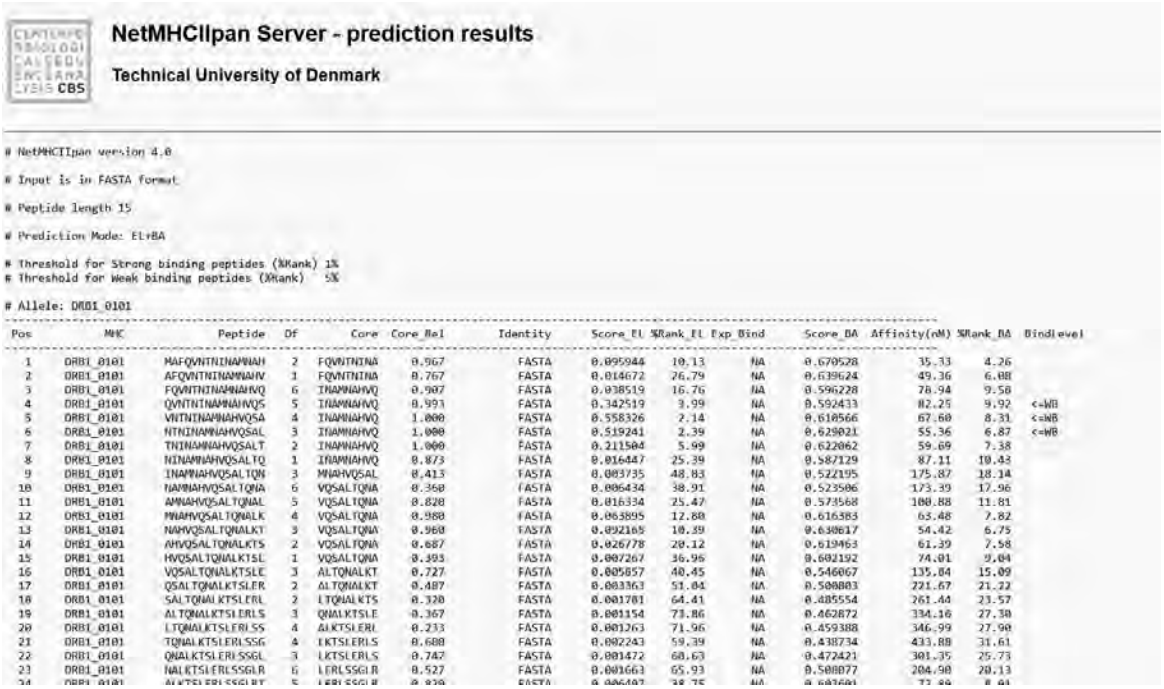
Figure 5:Toxicity prediction of CTL epitopes

Table 4:Final CTL epitopes selection

CTL	antigenicity	Allergenicity	Toxicity
ASDSQHLGF	0.4209 (Probable NON-ANTIGEN).	Non-Allergen	Non toxin
TSETGVEAY	1.7103 (Probable ANTIGEN).	Allergen	Non toxin
NILAQSGSY	-0.0323 (Probable NON-ANTIGEN).	Allergen	Non toxin
DLTKGSTNY	0.4841 (Probable NON-ANTIGEN)	Non-Allergen	Non toxin
ASGDISLTF	1.5112 (Probable ANTIGEN).	Non-Allergen	Non toxin

3.5 HTL Epitopes Detection and Sorting

To detect HTL epitopes and alleles, NetMHCIIpan 4.0 software was being used. Epitopes that have strong binding alleles were being selected (Figure 6).



NetMHCIIpan Server - prediction results

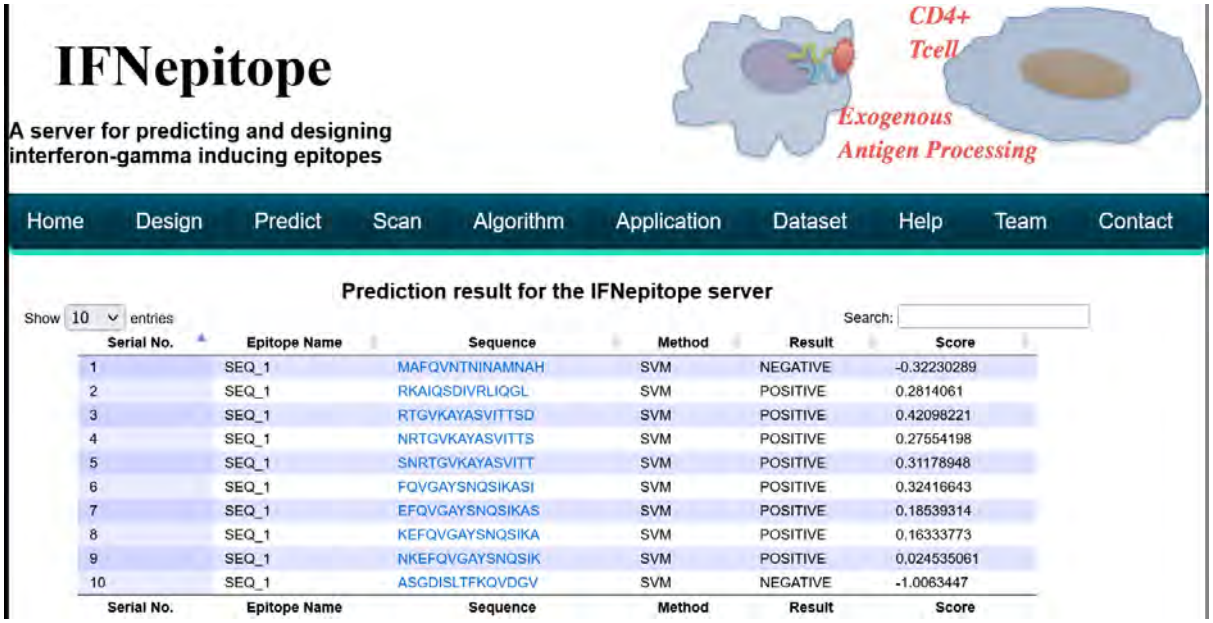
Technical University of Denmark

```
# NetMHCIIpan version 4.0

# Input is in FASTA format

# Peptide length 15

# Prediction Mode: EL+BA

# Threshold for Strong binding peptides (%Rank) 1%
# Threshold for Weak binding peptides (%Rank)   5%

# Allele: DRB1_0101
```

Pos	MHC	Peptide	Of	Core	Core_Rel	Identity	Score_EL	%Rank_EL	Exp_Bind	Score_BA	Affinity(nM)	%Rank_BA	BindLevel
1	DRB1_0101	MAFQVNTNINAMNAH	2	FQVNTNINA	0.967	FASTA	0.095944	10.13	NA	0.670528	35.33	4.26	
2	DRB1_0101	AFQVNTNINAMNAHV	1	FQVNTNINA	0.767	FASTA	0.014672	26.79	NA	0.639624	49.36	6.88	
3	DRB1_0101	FQVNTNINAMNAHVQ	6	INAMNAHVQ	0.907	FASTA	0.038519	16.76	NA	0.596228	78.94	9.50	
4	DRB1_0101	QVNTNINAMNAHVQS	5	INAMNAHVQ	0.993	FASTA	0.342519	3.99	NA	0.592433	82.25	9.92	<=WB
5	DRB1_0101	VNTNINAMNAHVQSA	4	INAMNAHVQ	1.000	FASTA	0.558326	2.14	NA	0.610566	67.60	8.31	<=WB
6	DRB1_0101	NTNINAMNAHVQSAL	3	INAMNAHVQ	1.000	FASTA	0.519241	2.39	NA	0.629021	55.36	6.87	<=WB
7	DRB1_0101	TNINAMNAHVQSALT	2	INAMNAHVQ	1.000	FASTA	0.211504	5.99	NA	0.622062	59.69	7.38	
8	DRB1_0101	NINAMNAHVQSALTQ	1	INAMNAHVQ	0.873	FASTA	0.016447	25.39	NA	0.587129	87.11	10.43	
9	DRB1_0101	INAMNAHVQSALTQN	3	MNAHVQSAL	0.413	FASTA	0.003735	48.83	NA	0.522195	175.87	18.14	
10	DRB1_0101	NAMNAHVQSALTQNA	6	VQSALTQNA	0.360	FASTA	0.006434	38.91	NA	0.523506	173.39	17.96	
11	DRB1_0101	AMNAHVQSALTQNAL	5	VQSALTQNA	0.828	FASTA	0.016334	25.47	NA	0.573568	100.88	11.81	
12	DRB1_0101	MNAHVQSALTQNALK	4	VQSALTQNA	0.980	FASTA	0.063895	12.80	NA	0.616383	63.48	7.82	
13	DRB1_0101	NAHVQSALTQNALKT	3	VQSALTQNA	0.960	FASTA	0.092165	10.39	NA	0.630617	54.42	6.75	
14	DRB1_0101	AHVQSALTQNALKTS	2	VQSALTQNA	0.887	FASTA	0.026778	20.12	NA	0.619463	61.39	7.58	
15	DRB1_0101	HVQSALTQNALKTSL	1	VQSALTQNA	0.393	FASTA	0.007267	36.96	NA	0.602192	74.01	9.04	
16	DRB1_0101	VQSALTQNALKTSLE	3	ALTQNALKT	0.727	FASTA	0.005857	40.45	NA	0.546067	135.84	15.09	
17	DRB1_0101	QSALTQNALKTSLER	2	ALTQNALKT	0.487	FASTA	0.003363	51.04	NA	0.500803	221.67	21.22	
18	DRB1_0101	SALTQNALKTSLERL	2	LTQNALKTS	0.320	FASTA	0.001701	64.41	NA	0.485554	261.44	23.57	
19	DRB1_0101	ALTQNALKTSLERLS	3	QNALKTSLE	0.367	FASTA	0.001154	73.86	NA	0.462872	334.16	27.30	
20	DRB1_0101	LTQNALKTSLERLSS	4	ALKTSLERL	0.233	FASTA	0.001263	71.96	NA	0.459388	346.99	27.90	
21	DRB1_0101	TQNALKTSLERLSSG	4	LKTSLERLS	0.680	FASTA	0.002243	59.39	NA	0.438734	433.88	31.61	
22	DRB1_0101	QNALKTSLERLSSGL	3	LKTSLERLS	0.747	FASTA	0.001472	68.63	NA	0.472421	301.35	25.73	
23	DRB1_0101	NALKTSLERLSSGLR	6	LERLSSGLR	0.527	FASTA	0.001663	65.93	NA	0.508077	204.90	20.13	
24	DRB1_0101	ALKTSLERLSSGLRT	5	LERLSSGLR	0.820	FASTA	0.006497	38.75	NA	0.603601	77.89	8.94	

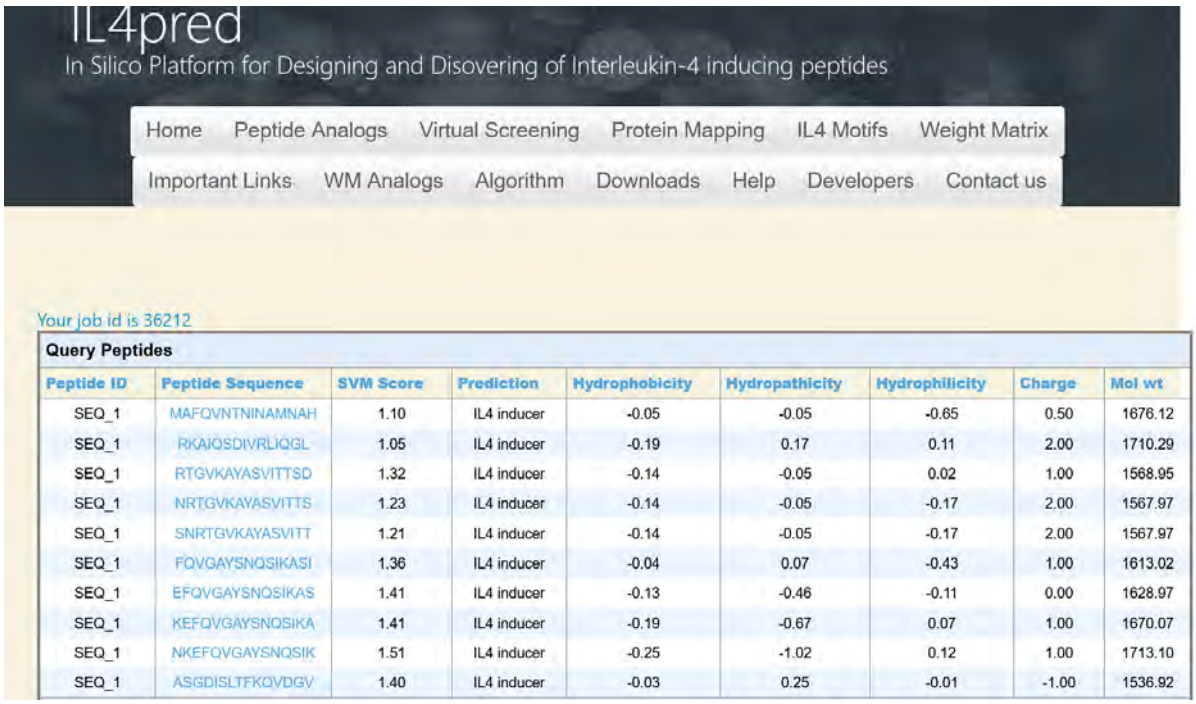
Figure 6:HTL epitopes detection for MHC class-II on NetMHCIIpan4.0 (Reynisson et al., 2020)

After that, sorting process were done through checking antigenicity, allergenicity, toxicity and cytokine inducing capacity of IFN-γ, IL-4, IL-10. Epitopes that passed through these criteria were being selected. IFN-γ inducing capacity was checked by IFNepitope software (Dhanda, Vir, et al., 2013); positive or not. Those showed positive result were selected (Figure 7). IL-4, IL-10 inducing capacity checked by IL4pred (Dhanda, Gupta, et al., 2013) and IL-10pred server (Nagpal et al., 2017). Those were showed result as inducer picked up (Figure 8,9). Among several epitopes two were selected (Table 5).



Serial No.	Epitope Name	Sequence	Method	Result	Score
1	SEQ_1	MAFQVNTNINAMNAH	SVM	NEGATIVE	-0.32230289
2	SEQ_1	RKAIQSDIVRLIQGL	SVM	POSITIVE	0.2814061
3	SEQ_1	RTGVKAYASVITTSD	SVM	POSITIVE	0.42098221
4	SEQ_1	NRTGVKAYASVITTS	SVM	POSITIVE	0.27554198
5	SEQ_1	SNRTGVKAYASVITT	SVM	POSITIVE	0.31178948
6	SEQ_1	FQVGAYSNQSIKASI	SVM	POSITIVE	0.32416643
7	SEQ_1	EFQVGAYSNQSIKAS	SVM	POSITIVE	0.18539314
8	SEQ_1	KEFQVGAYSNQSIKA	SVM	POSITIVE	0.16333773
9	SEQ_1	NKEFQVGAYSNQSIK	SVM	POSITIVE	0.024535061
10	SEQ_1	ASGDISLTFKQVDGV	SVM	NEGATIVE	-1.0063447
Serial No.	Epitope Name	Sequence	Method	Result	Score

Figure 7:IFNγ inducing abilities inspecting by IFNepitope server (Dhanda, Vir, et al., 2013)



Peptide ID	Peptide Sequence	SVM Score	Prediction	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Mol wt
SEQ_1	MAFQVNTNINAMNAH	1.10	IL4 inducer	-0.05	-0.05	-0.65	0.50	1676.12
SEQ_1	RKAIQSDIVRLIQGL	1.05	IL4 inducer	-0.19	0.17	0.11	2.00	1710.29
SEQ_1	RTGVKAYASVITTSD	1.32	IL4 inducer	-0.14	-0.05	0.02	1.00	1568.95
SEQ_1	NRTGVKAYASVITTS	1.23	IL4 inducer	-0.14	-0.05	-0.17	2.00	1567.97
SEQ_1	SNRTGVKAYASVITT	1.21	IL4 inducer	-0.14	-0.05	-0.17	2.00	1567.97
SEQ_1	FQVGAYSNQSIKASI	1.36	IL4 inducer	-0.04	0.07	-0.43	1.00	1613.02
SEQ_1	EFQVGAYSNQSIKAS	1.41	IL4 inducer	-0.13	-0.46	-0.11	0.00	1628.97
SEQ_1	KEFQVGAYSNQSIKA	1.41	IL4 inducer	-0.19	-0.67	0.07	1.00	1670.07
SEQ_1	NKEFQVGAYSNQSIK	1.51	IL4 inducer	-0.25	-1.02	0.12	1.00	1713.10
SEQ_1	ASGDISLTFKQVDGV	1.40	IL4 inducer	-0.03	0.25	-0.01	-1.00	1536.92

Figure 8: IL-4 inducing abilities inspecting on IL4pred server (Dhanda, Gupta, et al., 2013)

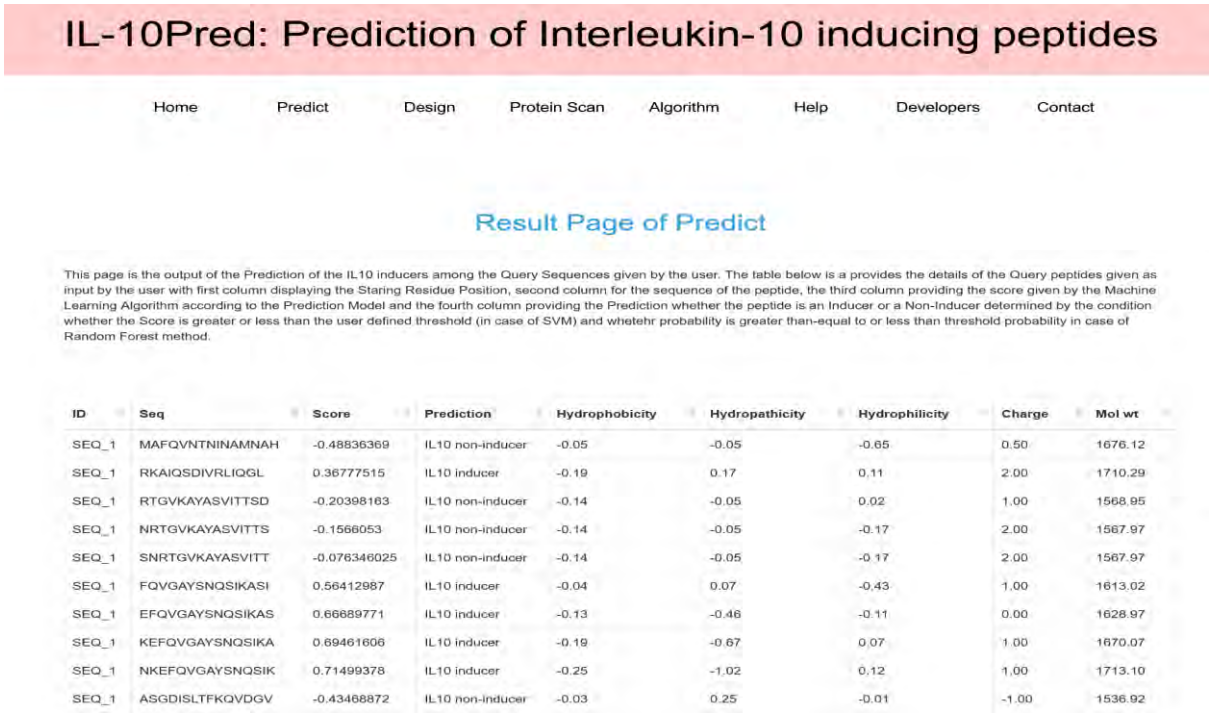


Figure 9: IL-10 inducing abilities inspecting on IL-10pred server (Nagpal et al., 2017)

Table 5:HTL Sorting with different parameters

HTL	Antigenicity	Allergenicity	IFNγ	IL-4	Il-10	Toxicity
AMDEQLKILDTVKVK	**0.4353**(Probable **NON-ANTIGEN**).	Allergen	negative	inducer	Non-inducer	Non-toxin
MDEQLKILDTVKVKA	**0.3734** (Probable **NON-ANTIGEN**).	Non-allergen	negative	Non inducer	Non inducer	Non toxin
DEQLKILDTVKVKAT	**0.5263** (Probable **ANTIGEN**)	Non-allergen	negative	Non inducer	Non inducer	Non toxin
EQLKILDTVKVKATQ	**0.4778** (Probable **NON-ANTIGEN**)	Non-allergen	negative	Non inducer	Non inducer	Non toxin
KIGQVRIATGALITA	**0.3170** (Probable **NON-ANTIGEN**).	allergen	positive	Non inducer	Non inducer	Non toxin
IGQVRIATGALITAS	**0.0489** (Probable **NON-ANTIGEN**).	Allergen	positive	Non inducer	Non inducer	Non toxin

GSGVTTLRGAMVVID	**0.8580** (Probable **ANTIGEN**)	Allergen	positive	Non inducer	Non inducer	Non toxin
NKNNILAQSGSYAMS	**0.0325** (Probable **NON-ANTIGEN**)	Allergen	negative	Inducer	inducer	Non toxin
KNNILAQSGSYAMSQ	**0.0746** (Probable **NON-ANTIGEN**).	Allergen	negative	Non inducer	Non inducer	Non toxin
AQSGSYAMSQANTVQ	**0.6325** (Probable **ANTIGEN**)	Non-allergen	positive	Non inducer	Non inducer	Non toxin
QSGSYAMSQANTVQQ	**0.5716** (Probable **ANTIGEN**).	Non-allergen	negative	Non inducer	Non inducer	Non toxin
SGSYAMSQANTVQQN	**0.7092** (Probable **ANTIGEN**).	Non-allergen	negative	Non inducer	Non inducer	Non toxin
GSYAMSQANTVQQNI	**0.5975** (Probable **ANTIGEN**)	Non-allergen	positive	Non inducer	Non inducer	Non toxin
VNTNINAMNAHVQSA	**1.3659** (Probable **ANTIGEN**).	Allergen	negative	Non inducer	Non inducer	Non toxin
GQVRIATGALITASG	**0.2230** (Probable **NON-ANTIGEN**)	Allergen	positive	Non inducer	Non inducer	Non toxin
GRLVAAINAVTSETG	**0.4655** (Probable **NON-ANTIGEN**)	Allergen	negative	Non inducer	Inducer	Non toxin
INAVTSETGVEAYTD	**1.0574** (Probable **ANTIGEN**).	Allergen	negative	Inducer	Non inducer	Non toxin
RLNLRSIDGRGIEIK	**3.0876** (Probable **ANTIGEN**)	Non-allergen	negative	Inducer	Non inducer	Non toxin
GIGVLAEVINKNSNR	**0.5388** (Probable **ANTIGEN**).	Non-allergen	negative	Non inducer	Non inducer	Non toxin
SANFNKNNILAQSGS	**0.2720** (Probable **NON-ANTIGEN**).	Allergen	negative	Non inducer	Inducer	Non toxin
LRDVTGNFNANVKSA	**0.7270** (Probable **ANTIGEN**).	Non-allergen	negative	Inducer	Non inducer	Non toxin

AFQVNTNINAMNAHV	**1.3419** (Probable **ANTIGEN**).	Allergen	negative	Non inducer	Non inducer	Non toxin
MAFQVNTNINAMNAH	**1.3022** (Probable **ANTIGEN**).	Allergen	negative	Non inducer	Non inducer	Non toxin
RKAIQSDIVRLIQGL	**-0.3348** (Probable **NON-ANTIGEN**).	Allergen	positive	Inducer	Inducer	Non toxin
RTGVKAYASVITTSD	**0.5942** (Probable **ANTIGEN**).	Allergen	positive	Inducer	Non inducer	Non toxin
NRTGVKAYASVITTS	**0.6905** (Probable **ANTIGEN**).	Non-allergen	positive	Inducer	Non inducer	Non toxin
SNRTGVKAYASVITT	**0.8208** (Probable **ANTIGEN**).	Non-allergen	positive	Inducer	Non inducer	Non toxin
FQVGAYSNQSIKASI	**0.8289** (Probable **ANTIGEN**)	Non-allergen	positive	Inducer	Inducer	Non toxin
EFQVGAYSNQSIKAS	**1.0776** (Probable **ANTIGEN**).	Non-allergen	positive	Inducer	Inducer	Non toxin
KEFQVGAYSNQSIKA	**0.7842** (Probable **ANTIGEN**).	Allergen	negative	Inducer	Inducer	Non toxin
NKEFQVGAYSNQSIK	**0.7497** (Probable **ANTIGEN**).	Allergen	positive	Inducer	Inducer	Non toxin
ASGDISLTFKQVDGV	**1.0824** (Probable **ANTIGEN**).	Non-allergen	negative	Inducer	Non inducer	Non toxin
TASGDISLTFKQVDG	**1.3780** (Probable **ANTIGEN**).	Allergen	negative	Inducer	Non inducer	Non toxin
TGNFNANVKSASGAN	**1.3244** (Probable **ANTIGEN**)	Non-allergen	negative	Inducer	Non inducer	Non toxin
VTGNFNANVKSASGA	**1.2665** (Probable **ANTIGEN**).	Allergen	negative	Inducer	Non inducer	Non toxin
DVTGNFNANVKSASG	**1.3252** (Probable **ANTIGEN**).	Non-allergen	negative	Inducer	Non inducer	Non toxin
RLSSGLRINKAADDA	**1.0267** (Probable **ANTIGEN**).	Non-allergen	negative	Non inducer	Non inducer	Non toxin

TVKVKATQAAQDGQT	**1.5166** (Probable **ANTIGEN**).	Allergen	negative	Inducer	Non inducer	Non toxin
LSSGLRINKAADDAS	**0.8770** (Probable **ANTIGEN**).	Allergen	negative	Non inducer	Non inducer	Non toxin
SGDISLTFKQVDGVN	**1.1421** (Probable **ANTIGEN**).	Non-allergen	negative	Inducer	Non inducer	Non toxin
KASIGSTTSDKIGQV	**1.1852** (Probable **ANTIGEN**).	Allergen	positive	Inducer	Non inducer	Non toxin
IKASIGSTTSDKIGQ	**1.3980** (Probable **ANTIGEN**).	Allergen	positive	Inducer	Non inducer	Non toxin
SIKASIGSTTSDKIG	**1.5371** (Probable **ANTIGEN**).	Allergen	positive	Inducer	Non inducer	Non toxin
TGVKAYASVITTSDV	**0.5517** (Probable **ANTIGEN**).	Non-allergen	positive	Inducer	Non inducer	Non toxin
VAAINAVTSETGVEA	**1.1629** (Probable **ANTIGEN**).	Allergen	negative	Inducer	Non inducer	Non toxin
LVAAINAVTSETGVE	**0.9892** (Probable **ANTIGEN**).	Allergen	negative	Non inducer	Non inducer	Non toxin
KAYASVITTSDVAVQ	**0.6566** (Probable **ANTIGEN**).	Allergen	positive	Inducer	Non inducer	Non toxin
VKAYASVITTSDVAV	**0.5635** (Probable **ANTIGEN**)	Allergen	positive	Inducer	Non inducer	Non toxin
GVKAYASVITTSDVA	**0.5348** (Probable **ANTIGEN**).	Non-allergen	positive	Inducer	Non inducer	Non toxin
GLRINKAADDASGMT	**1.4088** (Probable **ANTIGEN**)	Allergen	negative	Non inducer	Non inducer	Non toxin
SGLRINKAADDASGM	**1.1449** (Probable **ANTIGEN**).	Allergen	negative	Non inducer	Non inducer	Non toxin

SSGLRINKAADDASG	**1.5258** (Probable **ANTIGEN**).	Allergen	negative	Non inducer	Non inducer	Non toxin
NISITQVNVKAAESQ	**1.6160** (Probable **ANTIGEN**).	Allergen	negative	Inducer	Non inducer	Non toxin
SLGSGVTTLRGAMVV	**0.6584** (Probable **ANTIGEN**).	Allergen	positive	Non inducer	Non inducer	Non toxin
LGSGVTTLRGAMVVI	**0.5544** (Probable **ANTIGEN**).	Allergen	positive	Non inducer	Non inducer	Non toxin
GIEIKTDSVSNGPSA	**1.5218** (Probable **ANTIGEN**).	Allergen	negative	inducer	Non inducer	Non toxin
DASGMTVADSLRSQA	**0.9578** (Probable **ANTIGEN**).	Allergen	positive	Non inducer	Non inducer	Non toxin
ASGMTVADSLRSQAS	**0.8549** (Probable **ANTIGEN**).	Allergen	positive	Non inducer	Non inducer	Non toxin
VADSLRSQASSLGQA	**0.6334** (Probable **ANTIGEN**).	Allergen	positive	Non inducer	Non inducer	Non toxin
GRGIEIKTDSVSNGP	**1.6369** (Probable **ANTIGEN**).	Non-allergen	negative	Inducer	Non inducer	Non toxin
RGIEIKTDSVSNGPS	**1.4227** (Probable **ANTIGEN**).	Allergen	negative	Inducer	Non inducer	Non toxin
NTNINAMNAHVQSAL	**1.1124** (Probable **ANTIGEN**).	Non-allergen	negative	Non inducer	Non inducer	Non toxin
ADSLRSQASSLGQAI	**0.6902** (Probable **ANTIGEN**).	Non-allergen	positive	Non inducer	Non inducer	Non toxin
KILDTVKVKATQAAQ	**0.7148** (Probable **ANTIGEN**).	Non-allergen	negative	Non inducer	Non inducer	Non toxin
LDTVKVKATQAAQDG	**1.0971** (Probable **ANTIGEN**).	Allergen	negative	Inducer	Non inducer	Non toxin
DTVKVKATQAAQDGQ	**1.3118** (Probable **ANTIGEN**).	Non-allergen	negative	Inducer	Non inducer	Non toxin

LAEVINKNSNRTGVK	**1.0803** (Probable **ANTIGEN**).	Non-allergen	negative	Non inducer	Non inducer	Non toxin
AEVINKNSNRTGVKA	**1.3369** (Probable **ANTIGEN**).	Non-allergen	negative	Inducer	Inducer	Non toxin
TSETGVEAYTDQKGR	**1.5559** (Probable **ANTIGEN**).	Allergen	negative	Non inducer	Inducer	Non toxin
SETGVEAYTDQKGRL	**1.0660** (Probable **ANTIGEN**)	Non-allergen	negative	Non inducer	Inducer	Non toxin
ETGVEAYTDQKGRLN	**1.1432** (Probable **ANTIGEN**).	Non-allergen	negative	Non inducer	Inducer	Non toxin
DGRGIEIKTDSVSNG	**2.2120** (Probable **ANTIGEN**).	Non-allergen	negative	Inducer	Non inducer	Non toxin

3.6 Forecasting B-cell Epitopes

Bepipred linear epitope prediction 2.0 were being used for find out B cells epitopes with 0.5 threshold (Jespersen et al., 2017). Among several epitopes 6 were selected for final sorting They were sorted by antigenicity, allergenicity, toxicity (Table 6).

Table 6: Predicted B-cells and Sorting

B cells	Start	End	Length	Antigenicity	Allergenicity	toxicity
QNALKTSLERLSSGLRINKAADD	22	44	23	0.7116 (Probable ANTIGEN)	Allergen	Non toxin
RSQASSLGQAIA	55	66	12	0.5050 (Probable ANTIGEN).	Allergen	Non toxin
AAQDGQTTESRKAIQS	99	114	16	1.6745 (Probable ANTIGEN)	Non-Allergen	Non toxin
ISTVNNISITQVNVKAAESQIRDV	439	440	24	1.0090 (Probable ANTIGEN	Non-Allergen	Non toxin
KAMDEQLK	80	87	8	1.1889 (Probable ANTIGEN).	Allergen	Non toxin
TVQQNIL	500	506	7	0.9481 (Probable ANTIGEN).	Allergen	Non toxin

Server also provided a plotted graph with average, minimum and maximum number of B cell epitopes (Figure 10). The server gave output with 0.573 average, 0.233 minimum, 0.751 maximum score of B-cell epitopes.

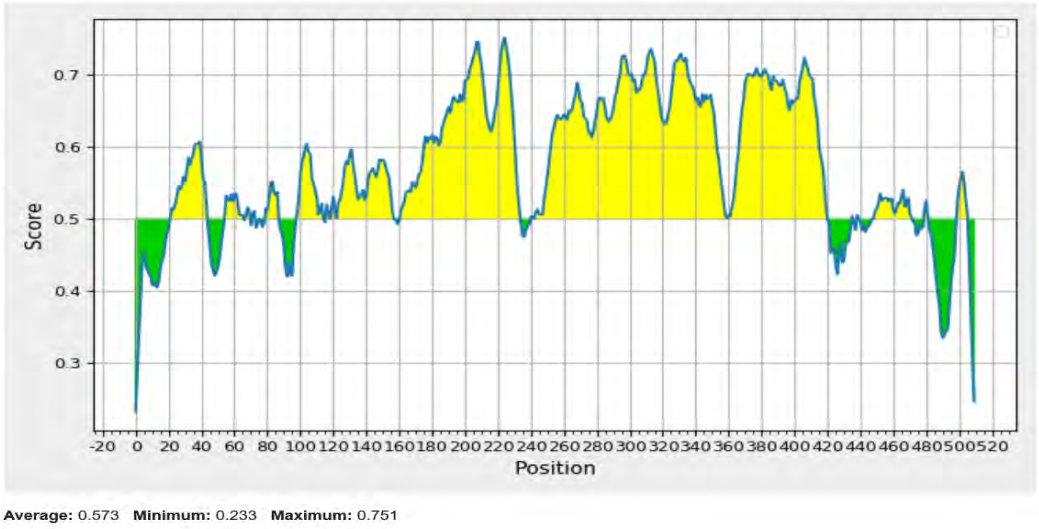


Figure 10: Score Vs position graph of B cell (Jespersen et al., 2017)

3.7 Composing of final vaccine

Final vaccine composition made with eligible epitopes that were being sorted with different parameters. The epitopes were added with adjuvant and for linking them linker EAAAK, GPGPG, KK were used. 1 CTL, 2 HTL, 2 B cell epitopes were added into sequence to adjuvant with EAAAK linker in the N-terminal location to elevate immunogenicity.

Constructed vaccine sequence:

MAFQVNTNINAMNAHVQSALTQNALKTSLER
LSSGLRINKAADDASGMTVADSLRSQASSLG
QAIANTNDGMGIIQVADKAMDEQLKILDTVK
VKATQAAQDGQTTESRKAIQSDIVRLIQGLD
NIGNTTTYNGQALLSGQFTNKEFQVGAYSNQ
SIKASIGSTTSDKIGQVRIATGALITASGDI
SLTFKQVDGVNDVTLESVKVSSSAGTGIGVL
AEVINKNSNRTGVKAYASVITTSDVAVQSGS
LSNLTLNGIHLGNIADIKKNDSDGRLVAAIN
AVTSETGVEAYTDQKGRLNLRSIDGRGIEIK
TDSVSNGPSALTMVNGGQDLTKGSTNYGRLS
LTRLDAKSINVVSASDSQHLGFTAIGFGESQ
VAETTVNLRDVTGNFNANVKSASGANYNAVI
ASGNQSLGSGVTTLRGAMVVIDIAESAMKML

DKVRSDLGSVQNQMISTVNNISITQVNVKAA
ESQIRDVDFAEESANFNKNNILAQSGSYAMS
QANTVQQNILRLLT**EAAAK**ASGDISLTF**GPG**
PGFQVGAYSNQSIKASI**GPGPG**EFQVGAYSN
QSIKAS**KK**AAQDGQTTESRKAIQS**KK**ISTVN
NISITQVNVKAAESQIRDV

3.8 Antigenicity of Constructed Vaccine

Vaccine's antigenicity again checked through Vaxijen v2.0 server. The antigenicity score came out better from the adjuvant's one which is a good indication of vaccine. The score is shown in figure 11.

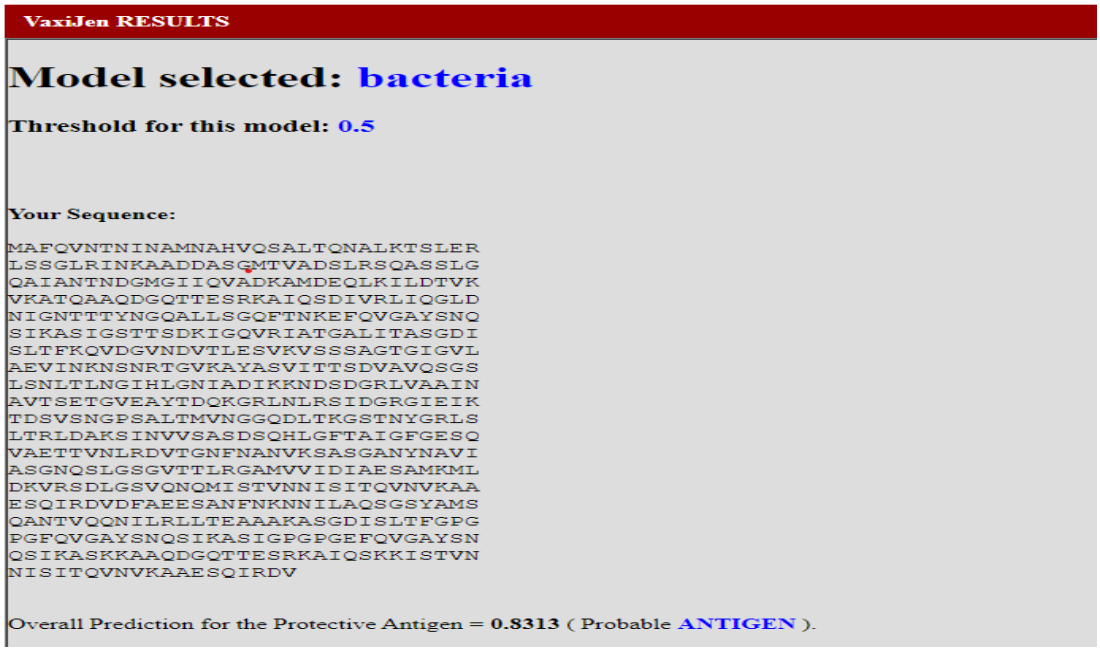


Figure 11: Antigenicity of Designed Vaccine (Doytchinova & Flower, 2007)

3.9 Evaluating Constructed vaccine's Biochemical feature

Protparam software was used to conduct biochemical feature in order to evaluate the vaccination. The Protparam server supply results of features e.g. molecular weight, molecular formula, instability index, GRAVY etc.

Number of amino acids, molecular weight, therapeutic pI are showed below (figure 12). The numbers got from results are eligible for vaccine.

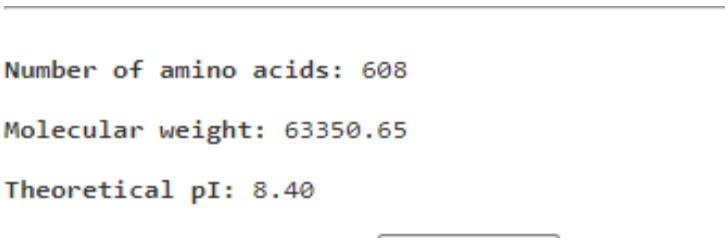


Figure 12: Vaccine's amino acid number, molecular weight, therapeutic pI feature on Protparam server (Gasteiger et al., n.d.)

Molecular formula, number of atoms, estimated half-life, instability index, aliphatic index, GRAVY results are shown below (Figure 13):

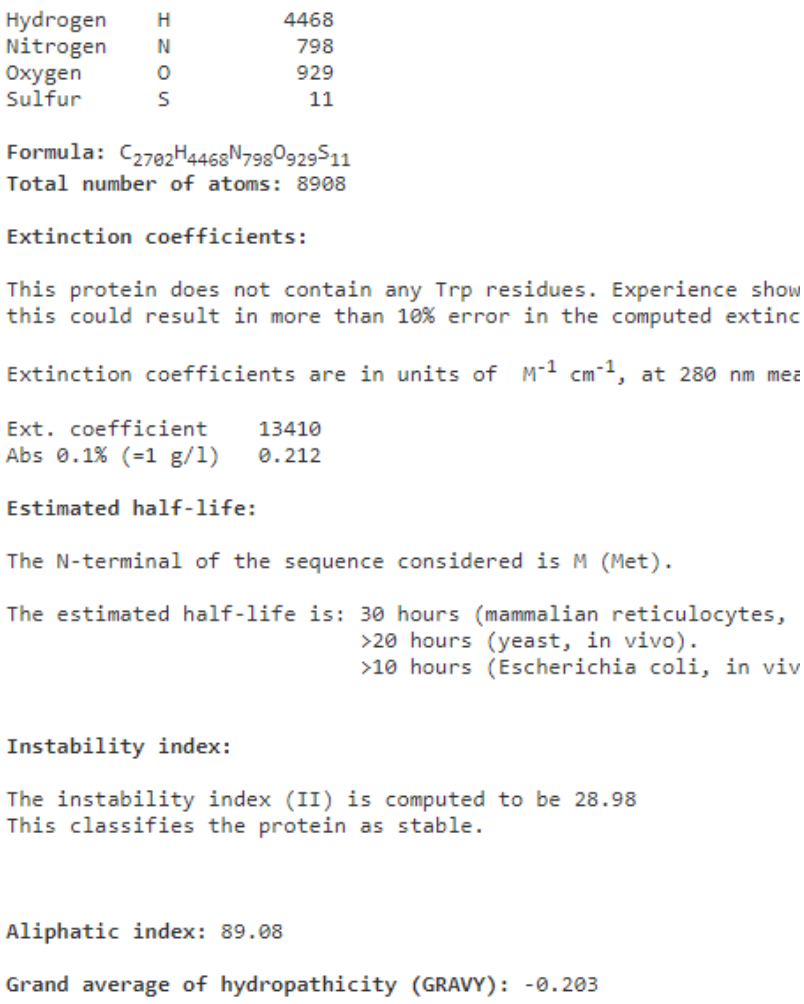


Figure 13: Vaccine's feature evaluation through estimated half-life, instability index, aliphatic index, GRAVY(Gasteiger et al., n.d.)

Grand average of hydropathicity (GRAVY) came out in negative form that indicates the vaccine is hydrophilic in nature. Instability index range came out in < 40 range which is 28.98; the vaccine is showing stable properties.

In addition, negative and positive charge residues, composition of amino acid was highlighted that are shown in figure 14:

Amino acid composition: CSV format

Ala (A)	49	10.0%
Arg (R)	15	3.1%
Asn (N)	41	8.4%
Asp (D)	24	4.9%
Cys (C)	0	0.0%
Gln (Q)	30	6.1%
Glu (E)	16	3.3%
Gly (G)	48	9.8%
His (H)	2	0.4%
Ile (I)	37	7.6%
Leu (L)	30	6.1%
Lys (K)	27	5.5%
Met (M)	6	1.2%
Phe (F)	11	2.3%
Pro (P)	5	1.0%
Ser (S)	59	12.1%
Thr (T)	38	7.8%
Trp (W)	0	0.0%
Tyr (Y)	9	1.8%
Val (V)	41	8.4%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 40
Total number of positively charged residues (Arg + Lys): 42

Figure 14: amino acid composition, negative and positive charge residue details by Protparam sever (Gasteiger et al., n.d.)

3.10 Allergenicity and Toxicity Checking of Designed Vaccine

The designed vaccine didn't reveal any allergenicity on AllergenOnline server that was searched by sliding 80 mer window FASTA (Figure 15). The result specify that server inspect 80 possible amino-acids of proteins tone with server database to find out not less than 35% recognition.

80mer Sliding Window Search Results

Database	AllergenOnline Database v22 (May 25, 2023)
Input Query	>FASTA MAFQVNTNINAMNAHVQSALTQNALKTSLERLSSGLRINKAADDASGMTVADSLRSQASS LGQAIANTNDGMGIIQVADKAMDEQLKILDTVKVKATQAAQDGQTTESRKAIQSDIVRLI QGLDNIGNTTTYNGQALLSGQFTNKEFQVGAYSNQSIKASIGSTTSDKIGQVRIATGALI TASGDISLTFKQVDGVNDVTLESVKVSSSAGTGIGVLAEVINKNSNRTGVKAYASVITTS DVAVQSGSLSNLTLNGIHLGNIADIKKNDSDGRLVAAINAVTSETGVEAYTDQKGRLNLR SIDGRGIEIKTDSVSNGPSALTMVNGGQDLTKGSTNYGRLSLTRLDAKSINVVSASDSQH LGFTAIGFGESQVAETTVNLRDVTGNFNANVKSASGANYNAVIASGNQSLGSGVTTLRGA MVVIDIAESAMKMLDKVRSDLGSVQNQMISTVNNISITQVNVKAAESQIRDVDFAEESAN FNKNNILAQSGSYAMSQANTVQQNILRLLTEAAAKASGDISLTFGPGPGFQVGAYSNQSI KASIGPGPGEFQVGAYSNQSIKASkkAAQDGQTTESRKAIQSkkISTVNNISITQVNVKA AESQIRDV
Length	608
Number of 80 mers	529
Number of Sequences with hits	0

No Matches of Greater than 35% Identity Found

AllergenOnline Database v22 (May 25, 2023)

Figure 15: Allergenicity identification on AllergenOnline database of Designed Vaccine (Goodman et al., 2016)

Besides that, toxicity examined through T3DB server to find out whether any toxin metabolite or particle present or not and the result came out with no toxicity (figure 16).

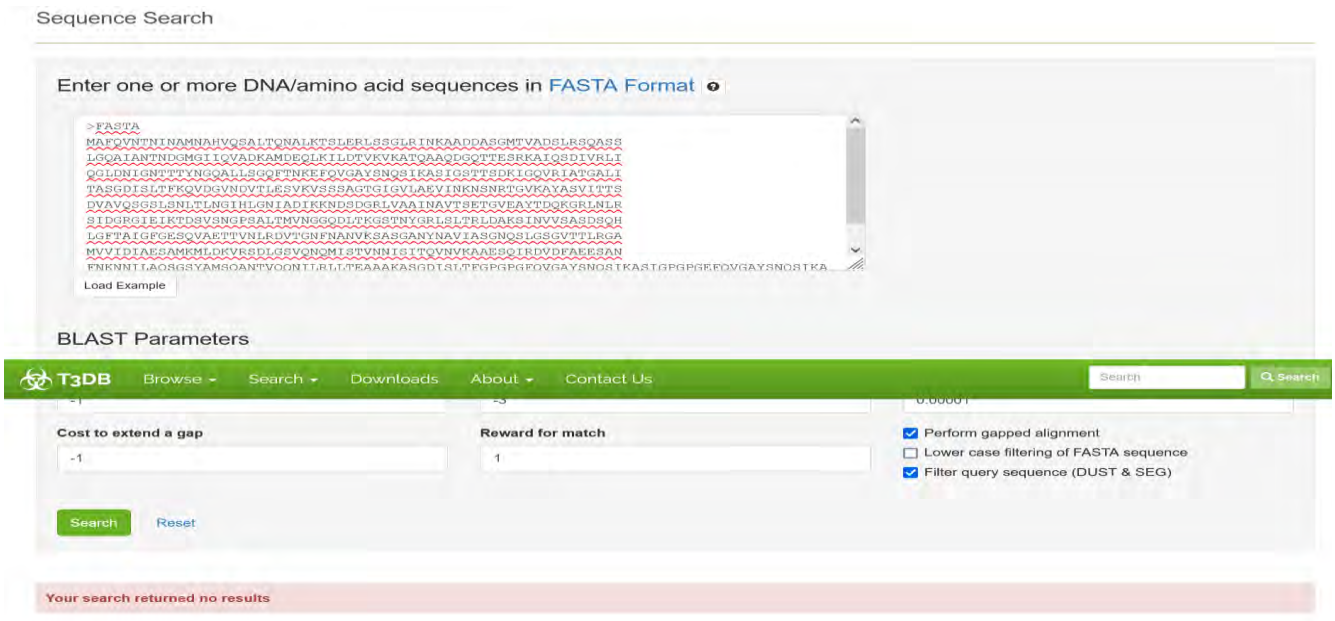


Figure 16: Toxicity identification on T3DB server of designed vaccine (Wishart et al., 2015)

3.11 Generation of 3D Model Designed Vaccine

Phyr^2 server (Kelley et al., 2015) helped to find out homology model of designed vaccine in the PDB file form. The confidence came out with 100% coverage and 83% coverage by working with 507 residues (Figure 17). The constructed 3D-model was opened on Discovery studio (Figure 18).

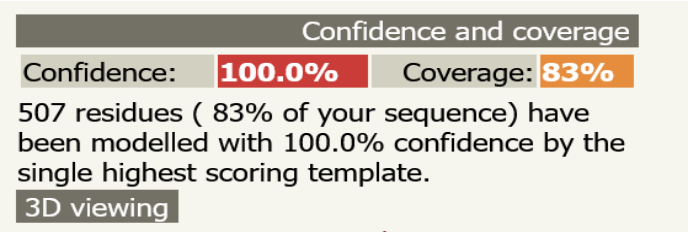


Figure 17: Confidence and coverage score from Phyr2 server (Kelley et al., 2015)

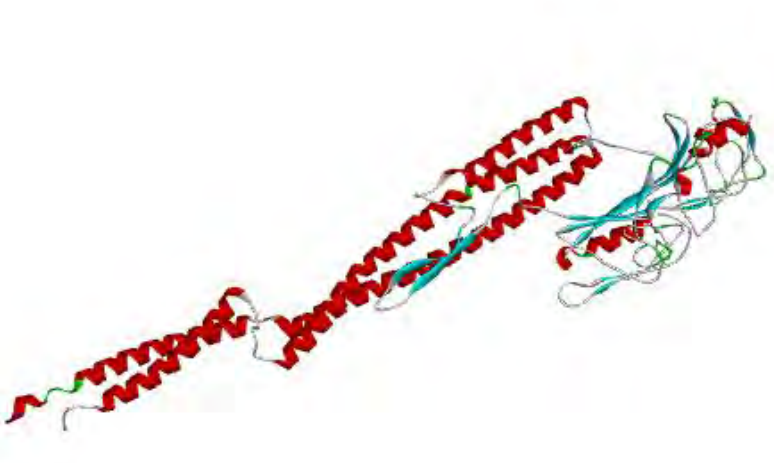


Figure 18: Constructed 3D model creation on phyr2 server (Kelley et al., 2015)

3.12 3D Model Validation Analysis

The PDB form file gotten from Phry^2 server used further on Swissmodel. Expasy server (Waterhouse et al., 2018) for validation with Ramachandran plot assessing (Figure 19). 88.87% region was showed as Ramachandran Favoured that specified that the mentioned amounts of

amino acids are situated in the that favoured region and 1.79% showed as Ramachandran outlier regions (Figure 20).

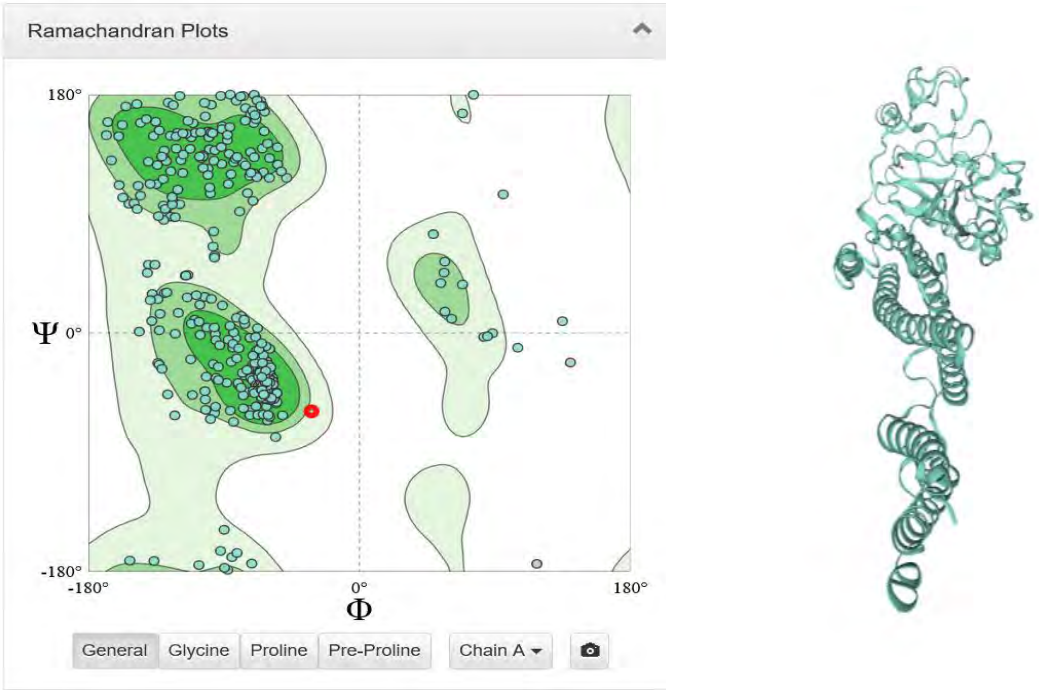


Figure 19: Ramachandran Plot on SWISS PDB plotter (Waterhouse et al., 2018)

MolProbity Results

	MolProbity Score	2.66	
☐	Clash Score	40.19	(A344 ARG-A352 VAL), (A365 ALA-A366 ILE), (A180 ILE-A351 ASN), (A180 ILE-A350 ILE), (A249 LEU-A314 VAL), (A315 SER-A316 ASN), (A330 LEU-
	Ramachandran Favoured	88.87%	
☐	Ramachandran Outliers	1.79%	A242 VAL, A207 SER, A316 ASN, A314 VAL, A391 VAL, A261 ASN, A257 ILE, A318 PRO, A366 ILE
	Rotamer Outliers	0.00%	
	C-Beta Deviations	0	
☐	Bad Bonds	6 / 3728	A15 HIS, A258 HIS, A360 HIS
☐	Bad Angles	5 / 5039	(A317 GLY-A318 PRO), (A5 VAL-A6 ASN), A15 HIS, A258 HIS, A360 HIS

Figure 20: Ramachandran plot's MolProbity results on SWISS PDB (Waterhouse et al., 2018)

3.13 3D Model Quality Analysis by Z-Score

The ProSA-web server provided model quality information through z-score Vs number of residue graph (Figure 21) and knowledge-based energy Vs sequence position graph (Figure 22). The z-score value -7.03 is in the native confirmation range. Structures from various origins (X-ray, NMR) are showed in distinct colors. In the second graph, the energy plot demonstrates local model quality by organizing energy as a function of amino acid sequence position and most of the residues are in the negative form excepts N-terminal region peaks which is slightly positive (Wiederstein & Sippl, 2007).

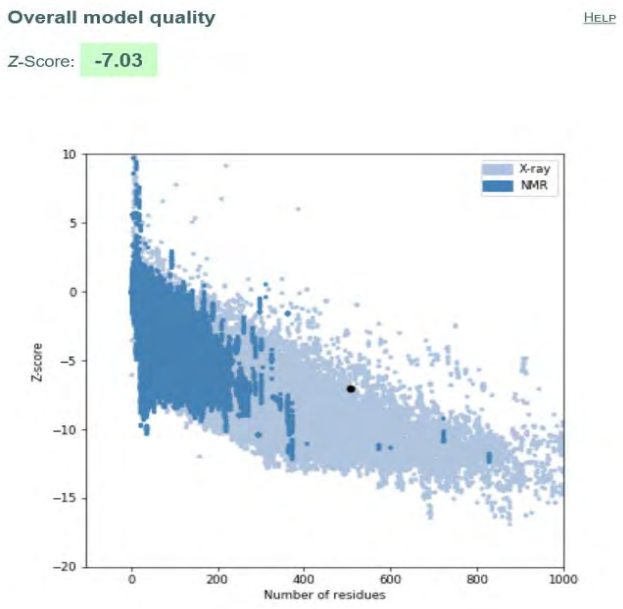


Figure 21: Z score Vs Number of residue graph on ProSA-web server (Wiederstein & Sippl, 2007)

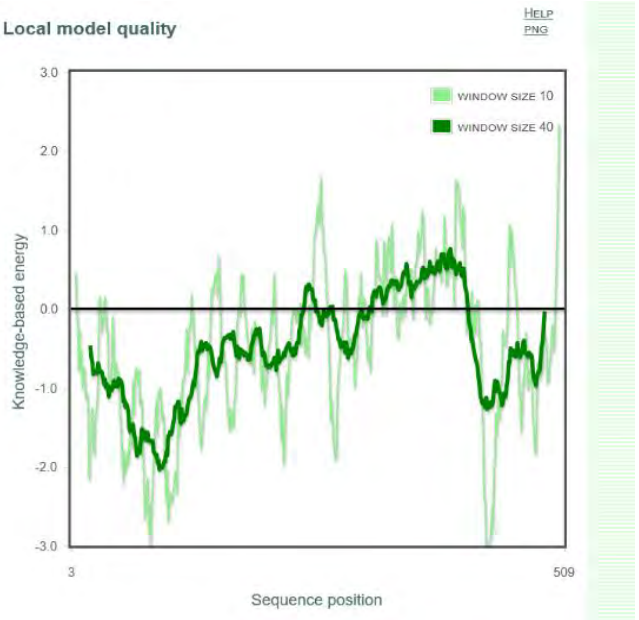


Figure 22: knowledge-based energy Vs sequence position graph on ProSA-web server (Wiederstein & Sippl, 2007)

3.14 Molecular Docking Assessing

The ClusPro server found out binding affinity between Human Toll like receptor (TLR5) and constructed vaccine to generate ligand-receptor complex. The server came out with results of 10 clusters among 30 clustered structures. Among them the cluster with lower negative score is considered as high ranked. High ranked cluster that shows interaction among TLR5 and designed vaccine has lowest ClusPro score of -1297.3. The ClusPro score represents sum of all the energies including van der Waals, electrostatic energies and more. The information of high ranked model showed in figure 23.

Cluster	Members	Representative	Weighted Score
0	43	Center	-1297.3
		Lowest Energy	-1297.3
1	42	Center	-1175.4
		Lowest Energy	-1175.4
2	42	Center	-1004.5
		Lowest Energy	-1193.5
3	30	Center	-1107.6
		Lowest Energy	-1140.8
4	29	Center	-1002.1
		Lowest Energy	-1289.4
5	23	Center	-1054.8
		Lowest Energy	-1171.6
6	22	Center	-1138.9
		Lowest Energy	-1138.9
7	22	Center	-1073.2
		Lowest Energy	-1105.4
8	22	Center	-1072.9
		Lowest Energy	-1134.9
9	22	Center	-1031.9
		Lowest Energy	-1101.4

Figure 23: Docking complex score of ClusPro server (Kozakov et al., n.d.)

The interactivity of designed vaccine and TLR5 that fetch from ClusPro sever is shown in Figure 24.

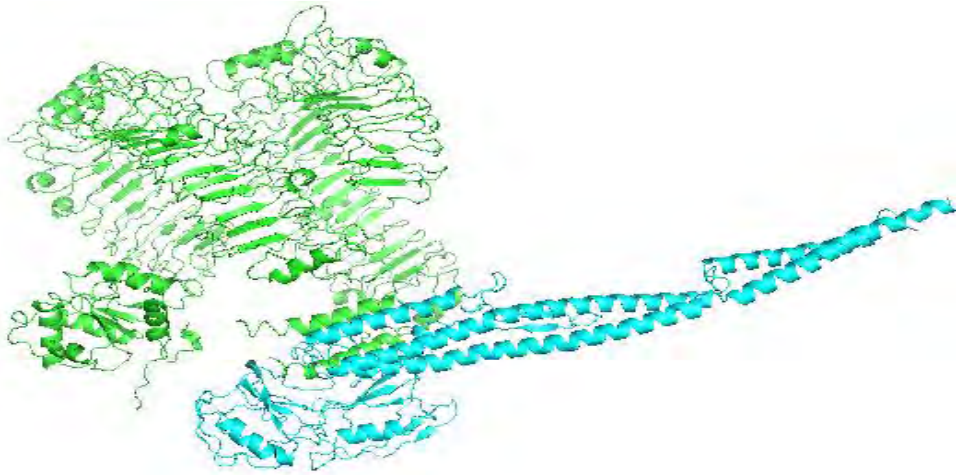
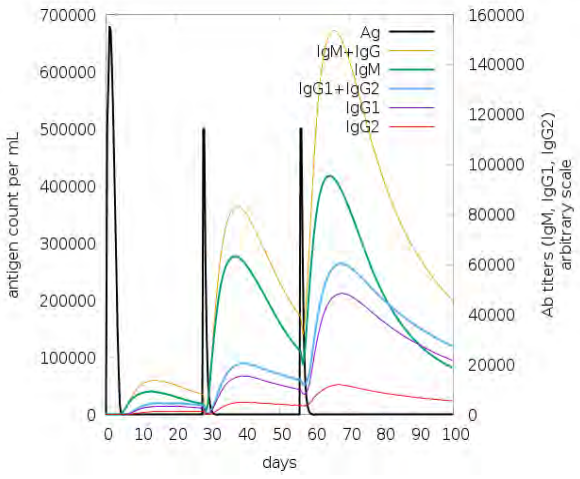


Figure 24: The interactivity of designed vaccine and TLR5 (Kozakov et al., n.d.)

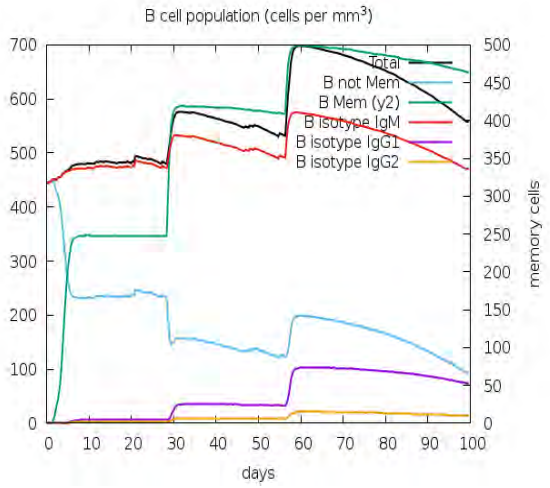
3.15 Vaccine's Immune Simulation

Immune simulation was carried out with C-immsim server of constructed vaccine to justify vaccine's immunoglobulin producing capacity with the respect of doses. The sever showed different graphica representation of inducing immune responses (Figure 25 A-J).

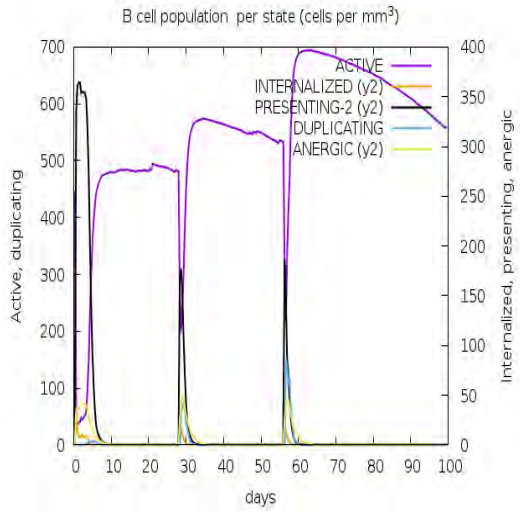


(A): Antibody titers and Antigen count graphical plot

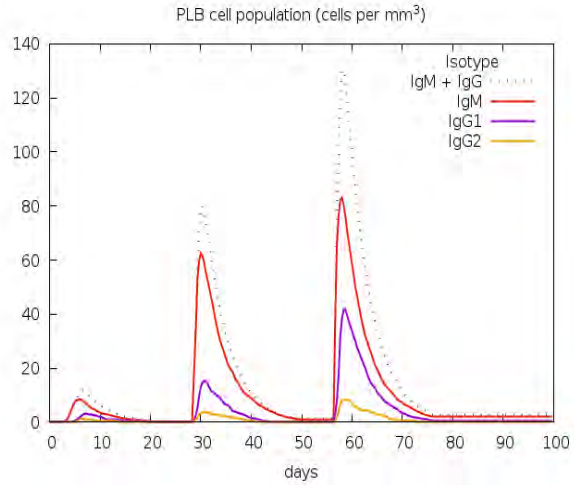
(Rapin et al., 2010)



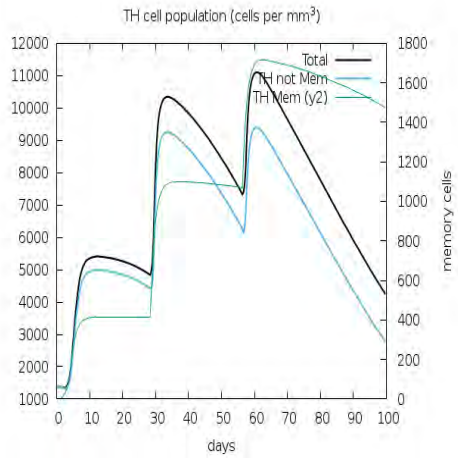
(B): B lymphocytes and memory cell's total count graphical plot (Rapin et al., 2010)



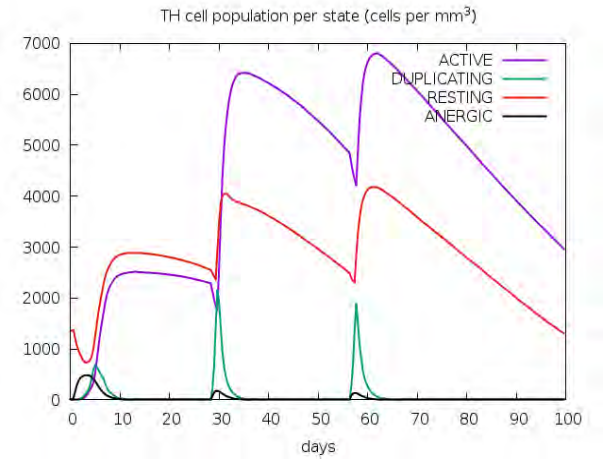
(C): B-cell population per entity state graphical plot (Rapin et al., 2010)



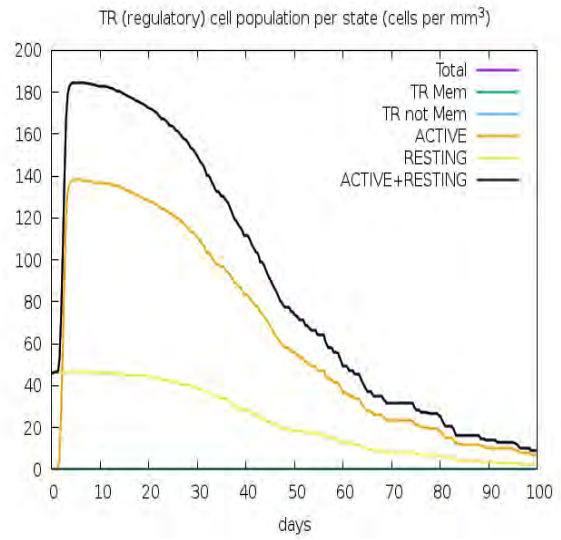
(D): Plasma B cell population based on its isotype graphical plot (Rapin et al., 2010)



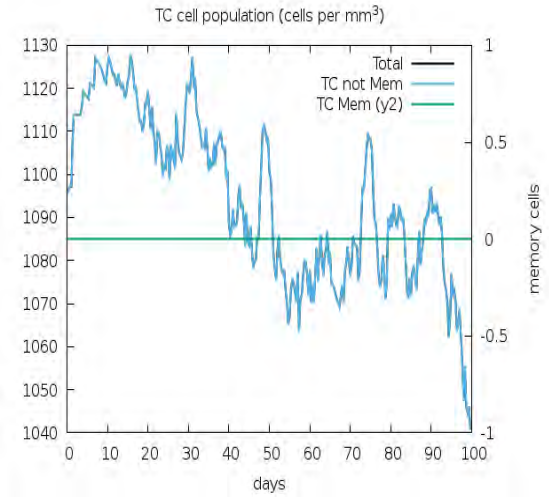
(E): Total and memory count of CD4 T- helper lymphocytes in graphical plot (Rapin et al., 2010)



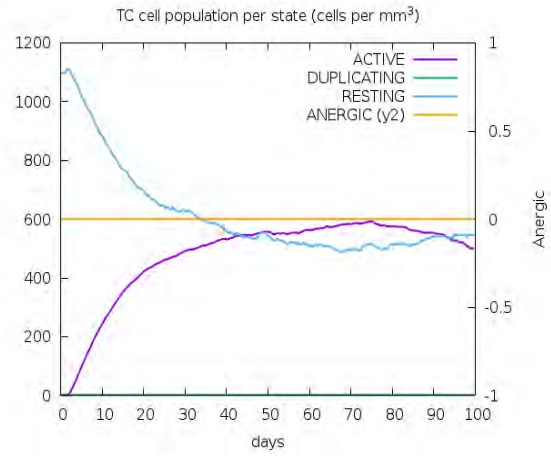
(F): CD4 T-Helper entity state in graphical plot (Rapin et al., 2010)



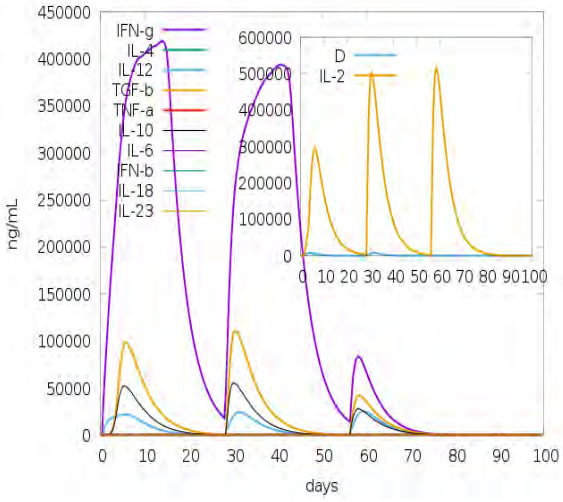
(G): Count of CD4 T-regulatory lymphocytes in graphical plot (Rapin et al., 2010)



(H): CD8 T cytotoxic lymphocyte count in graphical plot (Rapin et al., 2010)



(I): CD8 T-cytotoxic lymphocyte entity count in graphical plot (Rapin et al., 2010)



(J): Cytokines inducing levels in graphical plot (Rapin et al., 2010)

Figure 25(A-J): Graphical representation of immune-simulation on C-immsim (Rapin et al., 2010)

In figure (A) graphical representation, a rise of IgG and IgM antibodies were visible with antigen input in near 28 days. The rising peaks of antibodies were gradually increasing with

another dose input and almost reached in a topmost concentration after 60 days. Meanwhile, antigen peak was reduced little bit as it enhanced immune response. Figure (B) represent memory cell production from B lymphocytes cells that capture a memory of pathogen so that recognition will be occur in the future to identify the pathogen and will trigger immune response. Memory cells was gradually decrease after triggering antibody production form B cells. Figure (C) regard as B cell population entity state that visualize increase of activating and duplicating B cells number, while decreasing of anergic number. Plasma B cell population illustrated in figure (D) that showing gradually rising peaks of IgM, IgG1, IgG2 antibodies. Antibodies are relied on plasma B cells to mature and release them to play role in immune system. From figure (E), CD4 T helper (TH) cells increased with administered doses but after a while decrease level were noticed. Although CD4 TH cells level decreased a little bit, but not memory cells kept a reflection of antigen to use in future. Figure (F) showed entity state of CD4 TH cells that showed increase in activating, duplicating, resting cells and decrease in anergic cells of CD4 TH cells. Figure (G) represent that active regulatory CD4 cells were increased while vaccine doses administered and remain activated for almost 100 days. Furthermore, CD8 T cytotoxic (TC) lymphocytes count shown in figure (H); not memory cell were raised after administration, maintain rising position for several days and fell down in near 100 days. TC cells active, duplicating, resting, anergic cells number shown in figure (I). Lastly, figure (J) visualized different cytokine inducing levels including IFN-γ, IL-4, IL-10, IL-2 after applied doses.

Chapter 4

Discussion

Bangladesh has a high infection developing rate by *H. pylori* and they are associated with gastric cancer development in more than 80% cases as has been seen. Along with, *H. pylori* were linked with 660,000 cancer cases found in 2008 globally (Sarker et al., 2017). Excessive growth of *H. pylori* in the small intestine display signs as nausea, bloating, diarrhea, abdominal pain, flatulence signs and more (Dharan & Wozny, 2022). As consequences, a vaccine for *H. pylori* is highly needed. So, in-silico technique proposed a vaccine in this study. For that, a suitable protein flagellin (FlaA) was selected as it's antigenic part valuable for vaccine construction (Zarei et al., 2017).

The protein was gone through several steps to discover its CTL, HTL, B-cells epitopes with use of online prediction tools. After searching on NetCTL 1.2 server 6 CTL epitopes were gotten and sorting through the alleles on NetMHCIIpan4.0, settled on the best five CTL according to their strong binding. Once again, the epitopes sorted with different parameters (Toxicity, Antigenicity, Allergenicity) and this time an epitope was eligible.

A search using NetCTLIIpan 4.1 server almost yielded a total of 69 HTL, cytokine induing capabilities is used to categorize them with various internet server. 27 tested positive for IFN-γ, 37 was inducer of IL-4, 13 for IL-10. Again, epitopes sorted with antigenicity, allergenicity, toxicity and only 2 remained.

After using internet server Bedipred Liner Epitope Prediction 2.0 6 B cells were gotten and after further sorting two were chosen.

Sorted candidates used at multi epitope vaccine designing to combat *H. pylori* with primary protein and linkers. Vaccine's antigenicity has significantly increased compared to main protein. No allergic components, harmful entities were found from the projected experiment.

After making vaccine, an important factor is vaccine stability that conducted with ProtParam server and stability showed for final vaccine. Along that, preferred molecular weight, negative GRAVY score reflected eligibility of vaccine. Phry2 homology modelling reliably and comprehensively depicted the 3-Dimensional constitution with 100% coverage and 83% confidence. Positive findings were found in Ramachandran plot using Swiss.model expasy with favored region 88.87% and Z- score analysis using ProSA-web with -7.03 score of the vaccination. Binding of the constructed vaccine and toll like receptor 5 were displayed with ClusPro server score as -1297.3. Nonetheless, C-immsim server showed the anticipated response of antibody with higher IgG, IgM antibodies, CD4, CD8 cells after dosing input.

Chapter 5

Conclusion

To sum up, the research was about recommending a successful vaccination to oppose *H. pylori* bacteria since there is no viable vaccine in the market yet. Development of vaccine is motivated by antibiotic resistance and many diseases that are linked to the bacteria. Flagellin (FlaA) protein was targeted in order to create a multi epitope peptide vaccine. The key epitope selection and vaccine building was based on trustworthy computational technologies. The vaccine showed no toxin or allergic elements, satisfactory antigenicity, cytokine inducing capabilities, antibody producing potentiality, binding affinity of TLR5 and vaccine that might be able to generate simulation of immune responses. On the other hands still further in-vivo and in-vitro investigations are required to establish the quality, efficacy and safety of the developed vaccine.

References

Alexander, J., Fikes, J., Hoffman, S., Franke, E., Sacci, J., Appella, E., Chisari, F. V., Guidotti, L. G., Chesnut, R. W., Livingston, B., & Sette, A. (1998). The optimization of helper T lymphocyte (HTL) function in vaccine development. *Immunologic Research*, *18*.

Alm, R. A., Ling[2], L.-S. L., Moir[2], D. T., King, B. L., Brown, E. D., Doig, P. C., Smith[2], D. R., Noonan, B., Guild[2], B. C., Dejonge, B. L., Carmel[2], G., Tummino, P. J., Caruso[2], A., Uria-Nickelsen, M., Mills[2], D. M., Ives, C., Gibson[2], R., Merberg, D., Mills, S. D., … Trust, T. J. (1999). *Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori*.

Ayyagari, V. S., Venkateswarulu, T. C., Abraham Peele, K., & Srirama, K. (2022). Design of a multi-epitope-based vaccine targeting M-protein of SARS-CoV2: an immunoinformatics approach. *Journal of Biomolecular Structure & Dynamics*, *40*.

Bao, Y., Liu, X., Han, C., Xu, S., Xie, B., Zhang, Q., Gu, Y., Hou, J., Qian, L., Qian, C., Han, H., & Cao, X. (2014). Identification of IFN-γ-producing innate B cells. *Cell Research*, *24*.

Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bye-A-Jee, H., Cukura, A., Denny, P., Dogan, T., Ebenezer, T. G., Fan, J., Garmiri, P., da Costa Gonzales, L. J., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., … Zhang, J. (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*.

Camilo, V., Sugiyama, T., & Touati, E. (2017). Pathogenesis of Helicobacter pylori infection. *Helicobacter*, *22*.

Chehelgerdi, M., Heidarnia, F., Dehkordi, F. B., Chehelgerdi, M., Khayati, S., Khorramian-Ghahfarokhi, M., Kabiri-Samani, S., & Kabiri, H. (2023). Immunoinformatic prediction

of potential immunodominant epitopes from cagW in order to investigate protection against Helicobacter pylori infection based on experimental consequences. *Functional & Integrative Genomics*, *23*

Cid, T. P., Fernández, M. C., Benito Martínez, S., & Jones, N. L. (2013). Pathogenesis of Helicobacter pylori infection. *Helicobacter*, *18*

De Groot, A. S., Moise, L., Terry, F., Gutierrez, A. H., Hindocha, P., Richard, G., Hoft, D. F., Ross, T. M., Noe, A. R., Takahashi, Y., Kotraiah, V., Silk, S. E., Nielsen, C. M., Minassian, A. M., Ashfield, R., Ardito, M., Draper, S. J., & Martin, W. D. (2020). Better Epitope Discovery, Precision Immune Engineering, and Accelerated Vaccine Design Using Immunoinformatics Tools. *Frontiers in Immunology*, *11*.

Desta, I. T., Porter, K. A., Xia, B., Kozakov, D., & Correspondence, S. V. (2020). *Performance and Its Limits in Rigid Body Protein-Protein Docking*.

Dhanda, S. K., Gupta, S., Vir, P., & Raghava, G. P. (2013). Prediction of IL4 inducing peptides. *Clinical & Developmental Immunology*, *2013*, 263952.

Dhanda, S. K., Vir, P., & Raghava, G. P. S. (2013). Designing of interferon-gamma inducing MHC class-II binders. *Biology Direct*, *8*

Dharan, M., & Wozny, D. (2022). Helicobacter pylori infection and small intestinal bacterial overgrowth–more than what meets the eye. *World Journal of Clinical Cases*, *10*

Dimitrov, I., Flower, D. R., & Doytchinova, I. (2013). AllerTOP--a server for in silico prediction of allergens. *BMC Bioinformatics*, *14*

Ding, S. Z. (2020). Global whole family based- Helicobacter pylori eradication strategy to prevent its related diseases and gastric cancer. *World Journal of Gastroenterology*, *26*

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, *8*.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (n.d.). *Protein Analysis Tools on the ExPASy Server 571 571 From: The Proteomics Protocols Handbook Protein Identification and Analysis Tools on the ExPASy Server*.

Gong, W., Pan, C., Cheng, P., Wang, J., Zhao, G., & Wu, X. (2022). Peptide-Based Vaccines for Tuberculosis. *Frontiers in Immunology*, *13*.

Goodman, R. E., Ebisawa, M., Ferreira, F., Sampson, H. A., van Ree, R., Vieths, S., Baumert, J. L., Bohle, B., Lalithambika, S., Wise, J., & Taylor, S. L. (2016). AllergenOnline: A peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. *Molecular Nutrition & Food Research*, *60*

Goodwin, C. S., Armstrong, J. A., Chilvers, T., Peters, M., Collins, M. D., Sly, L., McConnell, W., & Harper, W. E. S. (1989). Transfer of Campylobacter pylori and Campylobacter mustelae to Helicobacter gen. nov. as Helicobacter pylori comb. nov. and Helicobacter mustelae comb. nov., respectively. *International Journal of Systematic Bacteriology*, *39*

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G. P. S. (2013). In silico approach for predicting toxicity of peptides and proteins. *PloS One*, *8*

HELICOBACTER PYLORI - Biological Agents - NCBI Bookshelf. (n.d.). Retrieved September 12, 2023, from https://www.ncbi.nlm.nih.gov/books/NBK304349/#a015.R200

Ito, H., & Seishima, M. (2010). Regulation of the Induction and Function of Cytotoxic T Lymphocytes by Natural Killer T Cell. *Journal of Biomedicine and Biotechnology*, *2010*.

Iyer, S. S., & Cheng, G. (2012). Role of Interleukin 10 Transcriptional Regulation in Inflammation and Autoimmune Disease. *Critical Reviews in Immunology*, *32*.

Jespersen, M. C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Research*, *45*

Kalkal, M., Kalkal, A., Dhanda, S. K., Das, E., Pande, V., & Das, J. (2022). A comprehensive study of epitopes and immune reactivity among Plasmodium species. *BMC Microbiology*, *22*

Kao, C. Y., Sheu, B. S., & Wu, J. J. (2016). Helicobacter pylori infection: An overview of bacterial virulence factors and pathogenesis. *Biomedical Journal*, *39*

Kar, P. P., Araveti, P. B., Kuriakose, A., & Srivastava, A. (2022). Design of a multi-epitope protein as a subunit vaccine against lumpy skin disease using an immunoinformatics approach. *Scientific Reports*, *12*

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*

Kira, J. ichi, & Isobe, N. (2019). Helicobacter pylori infection and demyelinating disease of the central nervous system. *Journal of Neuroimmunology*, *329*

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., & Vajda, S. (n.d.). *The ClusPro web server for protein-protein docking*.

Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Brunak, S., Lund, O., & Nielsen, M. (2005). An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *European Journal of Immunology*, *35*

Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., & Nielsen, M. (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics*, *8*.

Ma, J., Qiu, J., Wang, S., Ji, Q., Xu, D., Wang, H., Wu, Z., & Liu, Q. (2021). A Novel Design of Multi-epitope Vaccine Against Helicobacter pylori by Immunoinformatics Approach. *International Journal of Peptide Research and Therapeutics*, *27*

Matsuo, Y., Kido, Y., & Yamaoka, Y. (2017). Helicobacter pylori Outer Membrane Protein-Related Pathogenesis. *Toxins*, *9*

Montes-Grajales, D., & Olivero-Verbe, J. (2021). Bioinformatics Prediction of SARS-CoV-2 Epitopes as Vaccine Candidates for the Colombian Population. *Vaccines*, *9*

Nagpal, G., Usmani, S. S., Dhanda, S. K., Kaur, H., Singh, S., Sharma, M., & Raghava, G. P. S. (2017). Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Scientific Reports*, *7*

Nain, Z., Karim, M. M., Sen, M. K., & Adhikari, U. K. (2020). Structural basis and designing of peptide vaccine using PE-PGRS family protein of Mycobacterium ulcerans-An integrated vaccinomics approach. *Molecular Immunology*, *120*

None, M., P, B., None, P., M, S., & S, S. (2021). Prediction and identification of T cell epitopes of COVID-19 with balanced cytokine response for the development of peptide based vaccines. *In Silico Pharmacology*, *9*

Oh, J. D., Kling-Bä ckhed, H., Giannakis, M., Xu, J., Fulton, R. S., Fulton, L. A., Cordum, H. S., Wang, C., Elliott, G., Edwards, J., Mardis, E. R., Engstrand, L. G., & Gordon, J. I. (2006). *The complete genome sequence of a chronic atrophic gastritis Helicobacter pylori strain: Evolution during disease progression*.

Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational Immunology Meets Bioinformatics: The Use of Prediction Tools for Molecular Binding in the Simulation of the Immune System. *PLoS ONE*, *5*

Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Research*, *48*

Rueckert, C., & Guzmán, C. A. (2012). Vaccines: From Empirical Development to Rational Design. *PLoS Pathogens*, *8*

Sabbagh, P., Javanian, M., Koppolu, V., Vasigala, V. K. R., & Ebrahimpour, S. (2019). Helicobacter pylori infection in children: an overview of diagnostic methods. *European Journal of Clinical Microbiology & Infectious Diseases : Official Publication of the European Society of Clinical Microbiology*, *38*

Sarker, K. K., Kabir, Md. J., Bhuyian, A. K. M. M. uddin, Alam, Md. S., Chowdhury, F. R., Ahad, M. A., Rahman, Md. A., & Rahman, M. M. (2017). H. pylori infection and gastric cancer in Bangladesh: a case-control study. *International Journal of Surgery. Oncology*, *2*

Sircar, G., Sarkar, D., Bhattacharya, S. G., & Saha, S. (2014). Allergen databases. *Methods in Molecular Biology (Clifton, N.J.)*, *1184*

Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., … Craig Venter, J. (1997). The complete genome sequence of the gastric pathogen Helicobacter pylori. *Nature*, *388*

Tshibangu-Kabamba, E., & Yamaoka, Y. (2021). Helicobacter pylori infection and antibiotic resistance - from biology to clinical implications. *Nature Reviews. Gastroenterology & Hepatology*, *18*

Van Amsterdam, K., Van Vliet, A. H. M., Kusters, J. G., & Van Der Ende, A. (2006). Of microbe and man: Determinants of Helicobacter pylori-related diseases. *FEMS Microbiology Reviews*, *30*

Walker, F. C., Sridhar, P. R., & Baldridge, M. T. (2021). Differential roles of interferons in innate responses to mucosal viral infections. *Trends in Immunology*, *42*

Wang, H., Zhong, H., Gao, C., Zang, J., & Yang, D. (2021). The Distinct Properties of the Consecutive Disordered Regions Inside or Outside Protein Domains and Their Functional Significance. *International Journal of Molecular Sciences*, *22*

Wang, Y. C. (2014). Medicinal plant activity on Helicobacter pylori related diseases. *World Journal of Gastroenterology*, *20*

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*

Wishart, D., Arndt, D., Pon, A., Sajed, T., Guo, A. C., Djoumbou, Y., Knox, C., Wilson, M., Liang, Y., Grant, J., Liu, Y., Goldansaz, S. A., & Rappaport, S. M. (2015). T3DB: the toxic exposome database. *Nucleic Acids Research*, *43*

Zanotti, G., & Cendron, L. (2014). Structural and functional aspects of the Helicobacter pylori secretome. *World Journal of Gastroenterology : WJG*, *20*

Zarei, M., Mosayebi, G., Khansarinejad, B., & Abtahi, H. (2017). Antigenic and immunogenic evaluation of Helicobacter pylori FlaA epitopes. *Iranian Journal of Basic Medical Sciences*, *20*